Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order[*]

Radosław Adamczak[†], Paweł Wolff[†‡]

April 9, 2013

Abstract

Building on the inequalities for homogeneous tetrahedral polynomials in independent Gaussian variables due to R. Latała we provide a concentration inequality for non-necessarily Lipschitz functions $f\colon \mathbb{R}^n \to \mathbb{R}$ with bounded derivatives of higher orders, which hold when the underlying measure satisfies a family of Sobolev type inequalities

$$\|g - \mathbb{E}g\|_p \le C(p)\|\nabla g\|_p.$$

Such Sobolev type inequalities hold, e.g., if the underlying measure satisfies the log-Sobolev inequality (in which case $C(p) \le C\sqrt{p}$) or the Poincaré inequality (then $C(p) \le Cp$). Our concentration estimates are expressed in terms of tensor-product norms of the derivatives of f.

When the underlying measure is Gaussian and f is a polynomial (non-necessarily tetrahedral or homogeneous), our estimates can be reversed (up to a constant depending only on the degree of the polynomial). We also show that for polynomial functions, analogous estimates hold for arbitrary random vectors with independent sub-Gaussian coordinates.

We apply our inequalities to general additive functionals of random vectors (in particular linear eigenvalue statistics of random matrices) and the problem of counting cycles of fixed length in Erdős-Rényi random graphs, obtaining new estimates, optimal in a certain range of parameters.

Keywords: concentration of measure, Gaussian chaos, Sobolev inequalities
AMS Classification: Primary 60E15, 46N30; Secondary 60B20, 05C80

1   Introduction

Concentration of measure inequalities are one of the basic tools in modern probability theory (see the monograph [45]). The prototypic result for all concentration theorems is arguably the Gaussian concentration inequality [14, 60], which asserts that if G is a standard Gaussian vector in $\mathbb{R}^n$ and $f\colon \mathbb{R}^n \to \mathbb{R}$ is a 1-Lipschitz function, then for all $t > 0$,

$$\mathbb{P}(|f(G) - \mathbb{E}f(G)| \ge t) \le 2\exp(-t^2/2).$$

Over the years the above inequality has found numerous applications in the analysis of Gaussian processes, as well as in asymptotic geometric analysis (e.g. in modern proofs of Dvoretzky type theorems). Its applicability in geometric situations comes from the fact that it is dimension free and all norms in $\mathbb{R}^n$ are Lipschitz with respect to one another. However, there are some probabilistic or combinatorial situations, when one is concerned with functions that are

---

not Lipschitz. The most basic case is the probabilistic analysis of polynomials in independent random variables, which arise naturally, e.g., in the study of multiple stochastic integrals, in discrete harmonic analysis as elements of the Fourier expansions on the discrete cube or in numerous problems of random graph theory, to mention just the famous subgraph counting problem [35, 34, 22, 26, 25].

The concentration of measure or more generally integrability properties for polynomials have attracted a lot of attention in the last forty years. In particular Bonami [13] and Nelson [53] provided hypercontractive estimates (Khintchine type inequalities) for polynomials on the discrete cube and in the Gauss space, which have been later extended to other random variables by Kwapień and Szulga [40] (see also [41]). Khintchine type inequalities have been also obtained in the absence of independence for polynomials under log-concave measures by Bourgain [19], Bobkov [10], Nazarov-Sodin-Volberg [52] and Carbery-Wright [21].

Another line of research is to provide two sided estimates of moments of polynomials in terms of deterministic functions of the coefficients. Borell [15] and Arcones-Giné [5] provided such two sided bounds for homogeneous polynomials in Gaussian variables. They were expressed in terms of expectations of suprema of certain empirical processes. Talagrand [62] and Bousquet-Boucheron-Lugosi-Massart [18, 17] obtained counterparts of these results for homogeneous tetrahedral[1] polynomials in Rademacher variables and Łochowski [47] and Adamczak [1] for random variables with log-concave tails. Inequalities of this type, while implying (up to constants) hypercontractive bounds, have a serious downside as the analysis of the empirical processes involved is in general difficult. It is therefore important to obtain two-sided bounds in terms of purely deterministic quantities. Such bounds for random quadratic forms in independent symmetric random variables with log-concave tails have been obtained by Latała [42] (the case of linear forms was solved earlier by Gluskin and Kwapień in [28], whereas bounds for quadratic forms in Gaussian variables were obtained by Hanson-Wright [31], Borell [15] and Arcones-Giné [5]). Their counterparts for multilinear forms of arbitrary degree in nonnegative random variables with log-concave tails have been derived by Latała and Łochowski [44]. As for the symmetric case, the general problem is still open. An important breakthrough has been obtained by Latała [43], who proved two-sided estimates for Gaussian chaoses of arbitrary order, that is for homogeneous tetrahedral polynomials of arbitrary degree in independent Gaussian variables (we recall his bounds below as they are the starting point for our investigations). For general symmetric random variables with log-concave tails similar bounds are known only for chaoses of order at most three [2].

Polynomials in independent random variables have been also investigated in relation with combinatorial problems, e.g. with subgraph counting [35, 34, 22, 26, 25]. The best known result for general polynomial in this area has been obtained by Kim and Vu [36, 63], who presented a family of powerful inequalities for $[0, 1]$-valued random variables. Over the last decade they have been applied successfully to handle many problems in probabilistic combinatorics. Some recent inequalities for polynomials in the so called subexponential random variables have been also obtained by Schudy and Sviridenko [58, 57]. They are a generalization of the special case of exponential random variables in [44] and are expressed in terms of quantities similar to those considered by Kim-Vu.

Since it is beyond the scope of this paper to give a precise account of all the concentration inequalities for polynomials, we refer the reader to the aforementioned sources and recommend also the monographs [41, 23], where some parts of the theory are presented in a uniform way. As already mentioned we will present in detail only the results from [43], which are our main tool as well as motivation.

As for concentration results for general non-Lipschitz functions, the only reference we are aware of, which addresses this question is [29], where the Authors obtain interesting inequalities for stationary measures of certain Markov processes and functions satisfying a Lyapunov type

---

[1]A multivariate polynomial is called tetrahedral if all variables appear in it in power at most one.

condition. Their bounds are not comparable to the ones which we present in this paper. On the one hand they work in a more general Markov process setting, on the other hand, when specialized, e.g., to quadratic forms of Gaussian vectors, they do not recover optimal inequalities given in [15, 5, 43] (see Section 4 in [29]). Since the language of [29] is very different from ours, we will not describe the inequalities obtained therein and refer the interested reader to the original paper.

Let us now proceed to the presentation of our results. To do this we will first formulate a two sided tail and moment inequality for homogeneous tetrahedral polynomials in i.i.d. standard Gaussian variables due to Latała [43]. To present it in a concise way we need to introduce some notation which we will use throughout the article. For a positive integer n we will denote $[n] = \{1, \ldots, n\}$. The cardinality of a set I will be denoted by $\#I$. For $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ and $I \subseteq [d]$ we write $\mathbf{i}_I = (i_k)_{k \in I}$. We will also denote $|\mathbf{i}| = \max_{j \le d} i_j$.

Consider thus a d-indexed matrix $A = (a_{i_1, \ldots, i_d})_{i_1, \ldots, i_d = 1}^n$, such that $a_{i_1, \ldots, i_d} = 0$ whenever $i_j = i_k$ for some $j \ne k$, a sequence $g_1, \ldots, g_n$ of i.i.d. $\mathcal{N}(0, 1)$ random variables and define

$$Z = \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} g_{i_1} \cdots g_{i_d}. \tag{1}$$

Without loss of generality we can assume that the matrix A is symmetric, i.e., for all permutations $\sigma \colon [n] \to [n]$, $a_{i_1, \ldots, i_d} = a_{\sigma(i_1), \ldots, \sigma(i_d)}$.

Let now P_d be the set of partitions of $\{1, \ldots, d\}$ into nonempty, pairwise disjoint sets. For a partition $\mathcal{J} = \{J_1, \ldots, J_k\}$, and a d-indexed matrix $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ (non-necessarily symmetric or with zeros on the diagonal), define

$$\|A\|_{\mathcal{J}} = \sup \Big\{ \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} \prod_{l=1}^k x_{\mathbf{i}_{J_l}}^{(l)} \colon \|(x_{\mathbf{i}_{J_l}}^{(l)})\|_2 \le 1, 1 \le l \le k \Big\}, \tag{2}$$

where $\|(x_{\mathbf{i}_{J_l}})\|_2 = \sqrt{\sum_{|\mathbf{i}_{J_l}| \le n} x_{\mathbf{i}_{J_l}}^2}$. Thus, e.g.,

$$\|(a_{ij})_{i,j \le n}\|_{\{1,2\}} = \sup\{ \sum_{i,j \le n} a_{ij} x_{ij} \colon \sum_{i,j \le n} x_{ij}^2 \le 1\} = \sqrt{\sum_{i,j \le n} a_{ij}^2} = \|(a_{ij})_{i,j \le n}\|_{\mathrm{HS}},$$

$$\|(a_{ij})_{i,j \le n}\|_{\{1\}\{2\}} = \sup\{ \sum_{i,j \le n} a_{ij} x_i y_j \colon \sum_{i \le n} x_i^2 \le 1, \sum_{j \le n} y_j^2 \le 1\} = \|(a_{ij})_{i,j \le n}\|_{\ell_2^n \to \ell_2^n},$$

$$\|(a_{ijk})_{i,j,k \le n}\|_{\{1,2\}\{3\}} = \sup\{ \sum_{i,j,k \le n} a_{ijk} x_{ij} y_k \colon \sum_{i,j \le n} x_{ij}^2 \le 1, \sum_{k \le n} y_k^2 \le 1\}.$$

From the functional analytic perspective the above norms are injective tensor product norms of A seen as a multilinear form on $(\mathbb{R}^n)^d$ with the standard Euclidean structure.

We are now ready to present the inequalities by Latała. Below, as in the whole article by C_d we denote a constant, which depends only on d. The values of C_d may differ between occurrences.

Theorem 1.1. *For any d-indexed symmetric matrix $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ such that $a_{\mathbf{i}} = 0$ if $i_j = i_k$ for some $j \ne k$, the random variable Z, defined by (1) satisfies for all $p \ge 2$,*

$$C_d^{-1} \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A\|_{\mathcal{J}} \le \|Z\|_p \le C_d \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A\|_{\mathcal{J}}.$$

As a consequence, for all $t > 1$,

$$C_d^{-1} \exp\Big(-C_d \min_{\mathcal{J} \in P_d} \Big(\frac{t}{\|A\|_{\mathcal{J}}} \Big)^{2/\#\mathcal{J}} \Big) \le \mathbb{P}(|Z| \ge t) \le C_d \exp\Big(-\frac{1}{C_d} \min_{\mathcal{J} \in P_d} \Big(\frac{t}{\|A\|_{\mathcal{J}}} \Big)^{2/\#\mathcal{J}} \Big).$$

It is worthwhile noting that for $\#\mathcal{J} > 1$, the norms $\|A\|_{\mathcal{J}}$ are not unconditional in the standard basis (decreasing coefficients of the matrix may not result in decreasing the norm). Moreover, for specific matrices they may not be easy to compute. On the other hand, for any d-indexed matrix A and any $\mathcal{J} \in P_d$, we have $\|A\|_{\mathcal{J}} \leq \|A\|_{\{1,\ldots,d\}} = \sqrt{\sum_{\mathbf{i}} a_{\mathbf{i}}^2}$. Using this fact in the upper estimates above allows to recover (up to constants depending on d) hypercontractive estimates for homogeneous tetrahedral polynomials due to Nelson.

Our main result is an extension of the upper bound given in the above theorem to more general random functions and measures. Below we present the most basic setting we will work with and state the corresponding theorems. Some additional extensions are deferred to the main body of the article.

We will consider a random vector X in $\mathbb{R}^n$, which satisfies the following family of Sobolev inequalities. For any $p \geq 2$ and any smooth integrable function $f \colon \mathbb{R}^n \to \mathbb{R}$,

$$\|f(X) - \mathbb{E}f(X)\|_p \leq L\sqrt{p}\Big\||\nabla f(X)|\Big\|_p, \tag{3}$$

for some constant L (independent of p and f), where $|\cdot|$ is the standard Euclidean norm on $\mathbb{R}^n$. It is known (see [3] and Theorem 3.4 below) that if X satisfies the logarithmic Sobolev inequality with constant D_{LS}, then it satisfies (3) with $L = \sqrt{D_{LS}/2}$. We remark that there are many criteria for a random vector to satisfy the logarithmic Sobolev inequality (see e.g. [45, 7, 11, 8, 38]), so in particular our assumption (3) can be verified for many random vectors of interest.

Our first result is the following theorem, which provides moment estimates and concentration for D-times differentiable functions. The estimates are expressed by $\|\cdot\|_{\mathcal{J}}$ norms of derivatives of the function (which we will identify with multi-indexed matrices). We will denote the d-th derivative of f by $\mathbf{D}^d f$.

Theorem 1.2. *Assume that a random vector X in $\mathbb{R}^n$ satisfies the inequality* (3) *with constant L. Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function of the class $\mathcal{C}^D$. For all $p \geq 2$ if $\mathbf{D}^D f(X) \in L^p$, then*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C_D\Big(L^D \sum_{\mathcal{J} \in P_D} p^{\frac{\#\mathcal{J}}{2}}\Big\|\|\mathbf{D}^D f(X)\|_{\mathcal{J}}\Big\|_p + \sum_{1 \leq d \leq D-1} L^d \sum_{\mathcal{J} \in P_d} p^{\frac{\#\mathcal{J}}{2}}\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}\Big).$$

In particular if $\mathbf{D}^D f(x)$ is uniformly bounded on $\mathbb{R}^n$, then setting

$$\eta_f(t) = \min\left(\min_{\mathcal{J} \in P_D}\Big(\frac{t}{L^D \sup_{x \in \mathbb{R}^n}\|\mathbf{D}^D f(x)\|_{\mathcal{J}}}\Big)^{\frac{2}{\#\mathcal{J}}}, \min_{1 \leq d \leq D-1}\min_{\mathcal{J} \in P_d}\Big(\frac{t}{L^d\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}}\Big)^{\frac{2}{\#\mathcal{J}}}\right)$$

we obtain for $t > 0$,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp\Big(-\frac{1}{C_D}\eta_f(t)\Big).$$

The above theorem is quite technical, so we will now provide a few comments, comparing it to known results.

1. It is easy to see that if $D = 1$, Theorem 1.2 reduces (up to absolute constants) to the Gaussian-like concentration inequality, which can be obtained from (3) by Chebyshev's inequality (applied to general p and optimized).

2. If f is a homogeneous tetrahedral polynomial of degree D, then the tail and moment estimates of Theorem 1.2 coincide with those from Latała's Theorem. Thus Theorem 1.2 provides an extension of the upper bound from Latała 's result to a larger class of measures and functions (however we would like to stress that our proof relies heavily on Latała's work).

3. If f is a general polynomial of degree D, then $\mathbf{D}^D f(x)$ is constant on $\mathbb{R}^n$ (and thus equal to $\mathbb{E}\mathbf{D}^D f(X)$). Therefore in this case the function η_f appearing in Theorem 1.2 can be written in a simplified form

$$\eta_f(t) = \min_{1\leq d\leq D} \min_{\mathcal{J}\in P_d} \left(\frac{t}{L^d\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}}\right)^{2/\#\mathcal{J}}. \tag{4}$$

4. For polynomials in Gaussian variables, the estimates given in Theorem 1.2 can be reversed, like in Theorem 1.1. More precisely we have the following theorem, which provides an extension of Theorem 1.1 to general polynomials.

Theorem 1.3. *If G is a standard Gaussian vector in $\mathbb{R}^n$ and $f\colon \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree D, then for all $p \geq 2$,*

$$C_D^{-1} \sum_{1\leq d\leq D}\sum_{\mathcal{J}\in P_d} p^{\frac{\#\mathcal{J}}{2}}\|\mathbb{E}\mathbf{D}^d f(G)\|_{\mathcal{J}} \leq \|f(G) - \mathbb{E}f(G)\|_p \leq C_D \sum_{1\leq d\leq D}\sum_{\mathcal{J}\in P_d} p^{\frac{\#\mathcal{J}}{2}}\|\mathbb{E}\mathbf{D}^d f(G)\|_{\mathcal{J}}.$$

Moreover for all $t > 0$,

$$\frac{1}{C_D}\exp\left(-C_D\eta_f(t)\right) \leq \mathbb{P}(|f(G) - \mathbb{E}f(G)| \geq t) \leq C_D\exp\left(-\frac{1}{C_D}\eta_f(t)\right),$$

where

$$\eta_f(t) = \min_{1\leq d\leq D} \min_{\mathcal{J}\in P_d} \left(\frac{t}{\|\mathbb{E}\mathbf{D}^d f(G)\|_{\mathcal{J}}}\right)^{2/\#\mathcal{J}}.$$

5. It is well known that concentration of measure for general Lipschitz functions fails e.g. on the discrete cube and one has to impose some additional convexity assumptions to get sub-Gaussian concentration [61]. It turns out that if we restrict to polynomials, estimates in the spirit of Theorems 1.1 and 1.2 still hold. To formulate our result in full generality recall the definition of the ψ_2 Orlicz norm of a random variable Y,

$$\|Y\|_{\psi_2} = \inf\left\{t > 0\colon \mathbb{E}\exp\left(\frac{Y^2}{t^2}\right) \leq 2\right\}.$$

By integration by parts and Chebyshev's inequality $\|Y\|_{\psi_2} < \infty$ is equivalent to a sub-Gaussian tail decay for Y. We have the following result for polynomials in sub-Gaussian random vectors with independent components.

Theorem 1.4. *Let $X = (X_1, \ldots, X_n)$ be a random vector with independent components, such that for all $i \leq n$, $\|X_i\|_{\psi_2} \leq L$. Then for every polynomial $f\colon \mathbb{R}^n \to \mathbb{R}$ of degree D and every $p \geq 2$,*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C_D \sum_{d=1}^{D} L^d \sum_{\mathcal{J}\in \mathcal{P}_d} p^{\#\mathcal{J}/2}\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}.$$

As a consequence, for any $t > 0$,

$$\mathbb{P}\Big(|f(X) - \mathbb{E}f(X)| \geq t\Big) \leq 2\exp\left(-\frac{1}{C_D}\eta_f(t)\right),$$

where

$$\eta_f(t) = \min_{1\leq d\leq D} \min_{\mathcal{J}\in P_d} \left(\frac{t}{L^d\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}}\right)^{2/\#\mathcal{J}}.$$

6. We postpone the applications of our theorems to subsequent sections of the article and here we announce only that apart from polynomials we apply Theorem 1.2 to additive functionals and U-statistics of random vectors, in particular to linear eigenvalue statistics of random matrices, obtaining bounds which complement known estimates by Guionnet and Zeitouni [30]. Theorem 1.4 is applied to the special case of the problem of subgraph counting in large random graphs. In a special case when one counts copies of a given small cycle, our result allows to obtain optimal inequalities for random graphs $G(n, p)$, with $p \to 0$ slowly, namely $p \geq n^{-\frac{k-2}{2(k-1)}} \log^{-\frac{1}{2}} n$, where k is length of a cycle. To the best of our knowledge they are the best currently known inequalities for this range of p.

7. Let us now briefly discuss optimality of our inequalities. The lower bound in Theorem 1.3 clearly shows that Theorem 1.2 is optimal in the class of measures and functions it covers up to constants depending only on D. As for Theorem 1.4, it is similarly optimal in the class of random vectors with independent sub-Gaussian coordinates. In concrete combinatorial applications, for 0-1 random variables this theorem may be however suboptimal. This can be seen already for $D = 1$, for a linear combination of independent Bernoulli variables $X_1, \ldots, X_n$ with $\mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = p$. When p becomes small, the tail bound for such variables given e.g. by the Chernoff inequality is more subtle than what can be obtained from general inequalities for sums of sub-Gaussian random variables and the fact that $\|X_i\|_{\psi_2}$ is of order $(\log(2/p))^{-1/2}$. Roughly speaking, this is the reason why in our estimates for random graphs we have a restriction on the speed at which $p \to 0$. At the same time our inequalities still give results comparable to what can be obtained from other general inequalities for polynomials. As already noted in the survey [35], bounds obtained from various general inequalities for the subgraph-counting problem, may not be directly comparable, i.e. those performing well in one case may exhibit worse performance in some other cases. Similarly, our inequalities cannot be in general compared e.g. to the estimates by Kim and Vu. For this reason and since it would require introducing new notation, we will not discuss these inequalities and just indicate, when presenting applications of Theorem 1.4, several situations when our inequalities perform in a better or worse way than those by Kim and Vu. Let us only mention that the Kim-Vu inequalities similarly as ours are expressed in terms of higher order derivatives of the polynomials. However, Kim and Vu (as well as Schudy and Sviridenko) look at maxima of absolute values of partial derivatives, which does not lead to tensor-product norms which we consider. While in the general sub-Gaussian case we consider, such tensor product norms cannot be avoided (in view of Theorem 1.3), it is not necessarily the case for 0-1 random variables.

The organization of the paper is as follows. First, in Section 2, we introduce the notation used in the paper, next in Section 3 we give the proof of Theorem 1.2 together with some generalizations and examples of applications. In Section 4 we prove Theorem 1.3, whereas in Section 5 we present the proof of Theorem 1.4 and applications to the subgraph counting problems. In Section 6 we provide further refinements of estimates from Section 3 in the case of independent random variables satisfying modified log-Sobolev inequalities (they are deferred to the end of the article as they are more technical than those of Section 3). In the Appendix we collect some additional facts used in the proofs.

2 Notation

Sets and indices For a positive integer n we will denote $[n] = \{1, \ldots, n\}$. The cardinality of a set I will be denoted by $\#I$.

For $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ and $I \subseteq [d]$ we write $\mathbf{i}_I = (i_k)_{k \in I}$. We will also denote $|\mathbf{i}| = \max_{j \leq d} i_j$.

For a finite set A and an integer $d \geq 0$ we set

$$A^{\underline{d}} = \{\mathbf{i} = (i_1, \ldots, i_d) \in A^d \colon \forall_{j,k \in \{1,\ldots,d\}} \ j \neq k \Rightarrow i_j \neq i_k\}$$

(i.e. $A^{\underline{d}}$ is the set of d-indices with pairwise distinct coordinates). Accordingly we will denote $n^{\underline{d}} = n(n-1) \cdots (n-d+1)$.

By P_d we will denote the family of partitions of $[d]$ into nonempty, pairwise disjoint sets.

For a finite set I by $\ell_2(I)$ we will denote the finite dimensional Euclidean space $\mathbb{R}^I$ endowed with the standard Euclidean norm $|x|_2 = \sqrt{\sum_{i \in I} x_i^2}$. Whenever there is no risk of confusion we will denote the standard Euclidean norm simply by $|\cdot|$.

Multi-indexed matrices For a function $f \colon \mathbb{R}^n \to \mathbb{R}$ by $\mathbf{D}^d f(x)$ we will denote the (d-indexed) matrix of its derivatives of order d, which we will identify with the corresponding symmetric d-linear form. If $M = (M_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$, $N = (N_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ are d-indexed matrices, we define $\langle M, N \rangle = \sum_{\mathbf{i} \in [n]^d} M_{\mathbf{i}} N_{\mathbf{i}}$. Thus for all vectors $y_1, \ldots, y_d \in \mathbb{R}^n$ we have $\mathbf{D}^d f(x)(y_1, \ldots, y_d) = \langle \mathbf{D}^d f(x), y_1 \otimes \cdots \otimes y_d \rangle$, where $y_1 \otimes \cdots \otimes y_d = (y_{i_1} y_{i_2} \cdots y_{i_d})_{\mathbf{i} \in [n]^d}$.

We will also define the Hadamard product of two such matrices $M \circ N$ as a d-indexed matrix with entries $m_{\mathbf{i}} = M_{\mathbf{i}} N_{\mathbf{i}}$ (pointwise multiplication of entries).

Let us also define the notion of "generalized diagonals" of a d-indexed matrix $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$. For a fixed set $K \subseteq [d]$, with $\#K > 1$, the "generalized diagonal" corresponding to K is is the set of indices $\{\mathbf{i} \in [n]^d \colon i_k = i_l \text{ for } k, l \in K\}$.

Constants We will use the letter C to denote absolute constants and C_a for constants depending only on some parameter a. In both cases the values of such constants may differ between occurrences.

3 A concentration inequality for non-lipschitz functions

In this Section we prove Theorem 1.2. Let us first state our main tool, which is an inequality by Latała in a decoupled version.

Theorem 3.1 (Latała, [43])**.** *Let $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$ be a d-indexed matrix with real entries and let $G_1, G_2, \ldots, G_d$ be i.i.d. standard Gaussian vectors in $\mathbb{R}^n$. Let $Z = \langle A, G_1 \otimes \cdots \otimes G_d \rangle$. Then for every $p \geq 2$,*

$$C_d^{-1} \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A\|_{\mathcal{J}} \leq \|Z\|_p \leq C_d \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A\|_{\mathcal{J}}$$

Thanks to general decoupling inequalities for U-statistics [24], which we recall in the Appendix (Theorem 7.1), the above theorem is formally equivalent to Theorem 1.1. In fact in [43] Latała first proves the above version. In the proof of Theorem 3.3 we will need just Theorem 3.1 (in particular in this part of the article we do not need any decoupling inequalities).

From now on we will work in a more general setting than in Theorem 1.2 and assume that X is a random vector in $\mathbb{R}^n$, such that for all $p \geq 2$ there exists a constant $L_X(p)$ such that for all bounded $\mathcal{C}^1$ functions $f \colon \mathbb{R}^n \to \mathbb{R}$,

$$\|f(X) - \mathbb{E}f(X)\|_p \leq L_X(p) \big\| |\nabla f(X)| \big\|_p. \tag{5}$$

Clearly in this situation the above inequality generalizes to all $\mathcal{C}^1$ functions (if the right-hand side is finite then the left-hand side is well defined and the inequality holds).

Let now G be a standard n-dimensional Gaussian vector, independent of X. Using the Fubini theorem together with the fact that for some absolute constant C, all $x \in \mathbb{R}^n$ and $p \geq 2$, $C^{-1}\sqrt{p}|x| \leq \|\langle x, G\rangle\|_p \leq C\sqrt{p}|x|$, we can linearise the right-hand side above and write (5) equivalently (up to absolute constants) as

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \frac{CL_X(p)}{\sqrt{p}}\Big\|\langle \nabla f(X), G\rangle\Big\|_p. \tag{6}$$

We remark that similar linearisation has been used by Maurey and Pisier to provide a simple proof of the Gaussian concentration inequality [55, 56] (see remark following Theorem 3.3 below). Inequality (6) has an advantage over (5) as it allows for iteration leading to the following simple proposition.

Proposition 3.2. *Consider $p \geq 2$ and let X be an n-dimensional random vector satisfying* (5). *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^D$ function. Let moreover $G_1, \ldots, G_D$ be independent standard Gaussian vectors in $\mathbb{R}^n$, independent of X. Then for all $p \geq 2$, if $\mathbf{D}^D f(X) \in L^p$, then*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \frac{C^D L_X(p)^D}{p^{D/2}}\|\langle \mathbf{D}^D f(X), G_1 \otimes \cdots \otimes G_D\rangle\|_p \tag{7}$$
$$+ \sum_{1 \leq d \leq D-1} \frac{C^d L_X(p)^d}{p^{d/2}}\|\langle \mathbb{E}_X \mathbf{D}^d f(X), G_1 \otimes \cdots \otimes G_d\rangle\|_p.$$

Proof. Induction on D. For $D = 1$ the assertion of the proposition coincides with (6), which (as already noted) is equivalent to (5). Let us assume that the proposition holds for $D - 1$. Applying thus (7) with $D - 1$ instead of D, we obtain

$$\|f(X) - \mathbb{E}f(X)\|_p \leq \frac{C^{D-1} L_X(p)^{D-1}}{p^{(D-1)/2}}\|\langle \mathbf{D}^{D-1} f(X), G_1 \otimes \cdots \otimes G_{D-1}\rangle\|_p \tag{8}$$
$$+ \sum_{d=1}^{D-2} \frac{C^d L_X(p)^d}{p^{d/2}}\|\langle \mathbb{E}_X \mathbf{D}^d f(X), G_1 \otimes \cdots \otimes G_d\rangle\|_p.$$

Applying now the triangle inequality in L^p, we get

$$\|\langle \mathbf{D}^{D-1} f(X), G_1 \otimes \cdots \otimes G_{D-1}\rangle\|_p \leq \|\langle \mathbf{D}^{D-1} f(X) - \mathbb{E}_X \mathbf{D}^{D-1} f(X), G_1 \otimes \cdots \otimes G_{D-1}\rangle\|_p$$
$$+ \|\langle \mathbb{E}_X \mathbf{D}^{D-1} f(X), G_1 \otimes \cdots \otimes G_{D-1}\rangle\|_p. \tag{9}$$

Let us now apply (6) conditionally on $G_1, \ldots, G_{D-1}$ to the function $f_1(x) = \langle \mathbf{D}^{D-1} f(x), G_1 \otimes \cdots \otimes G_{D-1}\rangle$. Since $\langle \mathbf{D}^{D-1} f(X) - \mathbb{E}_X \mathbf{D}^{D-1} f(X), G_1 \otimes \cdots \otimes G_{D-1}\rangle = f_1(X) - \mathbb{E}_X f_1(X))$ and $\langle \nabla f_1(X), G_D\rangle = \langle \mathbf{D}^D f(X), G_1 \otimes \cdots \otimes G_D\rangle$, we obtain

$$\mathbb{E}_X |\langle \mathbf{D}^{D-1} f(X) - \mathbb{E}_X \mathbf{D}^{D-1} f(X), G_1 \otimes \cdots \otimes G_{D-1}\rangle|^p$$
$$\leq \frac{C^p L_X(p)^p}{p^{p/2}} \mathbb{E}_{X, G_D} |\langle \mathbf{D}^D f(X), G_1 \otimes \cdots \otimes G_D\rangle|^p.$$

To finish the proof it is now enough to integrate this inequality with respect to the remaining Gaussian vectors and combine the obtained estimate with (8) and (9). $\qquad \square$

Let us now specialize to the case when $L_X(p) = Lp^\gamma$ for some $L > 0, \gamma \geq 1/2$. Combining the above proposition with Latała's Theorem 3.1, we obtain immediately the following theorem, a special case of which is Theorem 1.2.

Theorem 3.3. *Assume that X is a random vector in $\mathbb{R}^n$, such that for some constants $L > 0, \gamma \geq 1/2$, all smooth functions f and all $p \geq 2$,*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq Lp^\gamma \Big\|\|\nabla f(X)\|\Big\|_p. \tag{10}$$

8

For any smooth function $f \colon \mathbb{R}^n \to \mathbb{R}$ of class $\mathcal{C}^D$ and $p \geq 2$ if $\mathbf{D}^D f(X) \in L^p$, then

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C_D \Big(\sum_{\mathcal{J} \in P_D} L^D p^{(\gamma - 1/2)D + \#\mathcal{J}/2} \Big\| \|\mathbf{D}^D f(X)\|_{\mathcal{J}} \Big\|_p$$

$$+ \sum_{1 \leq d \leq D-1} \sum_{\mathcal{J} \in P_d} L^d p^{(\gamma - 1/2)d + \#\mathcal{J}/2} \|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}} \Big).$$

If $\mathbf{D}^D f$ is bounded uniformly on $\mathbb{R}^n$, then for all $t > 0$,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2 \exp\Big(- \frac{1}{C_D} \eta_f(t) \Big),$$

where

$$\eta_f(t) = \min(A, B),$$
$$A = \min_{\mathcal{J} \in P_D} \left(\Big(\frac{t}{L^D \sup_{x \in \mathbb{R}^n} \|\mathbf{D}^D f(x)\|_{\mathcal{J}}} \Big)^{2/((2\gamma-1)D + \#\mathcal{J})} \right),$$
$$B = \min_{1 \leq d \leq D-1} \min_{\mathcal{J} \in P_d} \left(\Big(\frac{t}{L^d \|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}} \Big)^{2/((2\gamma-1)d + \#\mathcal{J})} \right).$$

Proof. The first part is a straightforward combination of Proposition 3.2 and Theorem 3.1. The second part follows from the first one by Chebyshev's inequality $\mathbb{P}(|Y| \geq e\|Y\|_p) \leq \exp(-p)$ applied with $p = \eta_f(t)/C_D$ (note that if $\eta_f(t)/C_D \leq 2$ then one can make the tail bound asserted in the theorem trivial by adjusting the constants). $\qquad\square$

Remark In [55, 56] Pisier presents a stronger inequality than (10) with $\gamma = 1/2$. More specifically, he proves that if X, G are independent standard centred Gaussian vectors in $\mathbb{R}^n$, E is a Banach space and $f \colon \mathbb{R}^n \to E$ is a $\mathcal{C}^1$ function, then for every convex function $\Phi \colon E \to \mathbb{R}$,

$$\mathbb{E}\Phi(f(X) - \mathbb{E}f(X)) \leq \mathbb{E}\Phi\Big(L\langle \nabla f(X), G \rangle \Big), \tag{11}$$

where $L = \frac{\pi}{2}$. As noted in [46], Caffarelli's contraction principle [20] implies that, e.g., a random vector X with density e^{-V}, where $V \colon \mathbb{R}^n \to \mathbb{R}$ satisfies $D^2 V \geq \lambda \mathrm{Id}$, $\lambda > 0$ satisfies the above inequality with $L = \frac{\pi}{2\sqrt{\lambda}}$ (where G is still a standard Gaussian vector independent of X). Therefore in this situation a similar approach as in the proof of Proposition 3.2 can be used for functions f with values in a general Banach space. Moreover, a counterpart of Latała's results is known for chaoses with values in a Hilbert space (to the best of our knowledge this observation has not been published, in fact it can be quite easily obtained from the version for real valued chaoses). Thus in this case we can obtain a counterpart of Theorem 3.3 (with $\gamma = 1/2$) for Hilbert space valued-functions. In the case of a general Banach space two-sided estimates for Banach space-valued Gaussian chaoses are not known. Still, one can use some known inequalities (like hypercontraction or Borell-Arcones-Giné inequality) instead of Theorem 3.1 and thus obtain new concentration bounds. We remark that if one uses hypercontraction, one can obtain explicit dependence of the constants on the degree of the polynomial, since explicit constants are known for hypercontractive estimates of (Banach space-valued) Gaussian chaoses and one can keep track of them during the proof. We skip the details.

In view of Theorem 3.3 a natural question arises: for what measures is the inequality (10) satisfied? Before we provide examples, for technical reasons let us recall the definition of the length of the gradient of a locally Lipschitz function. For a metric space $(\mathcal{X}, d)$, a locally Lipschitz function $f \colon \mathcal{X} \to \mathbb{R}$ and $x \in \mathcal{X}$, we define

$$|\nabla f|(x) = \limsup_{d(x,y) \to 0} \frac{|f(y) - f(x)|}{d(x, y)}. \tag{12}$$

If $\mathcal{X} = \mathbb{R}^n$ and f is differentiable at x, then clearly $|\nabla f|(x)$ coincides with the Euclidean length of the usual gradient $\nabla f(x)$. For this reason, with slight abuse of notation, we will write $|\nabla f(x)|$ instead of $|\nabla f|(x)$. We will consider only measures on $\mathbb{R}^n$, however since we allow measures which are not necessarily absolutely continuous with respect to the Lebesgue measure, at some points in the proofs we will work with the above abstract definition.

Going back to the question of measures satisfying (10), it is well known (see e.g. [50]) that if X satisfies the Poincaré inequality

$$\mathrm{Var}\,(f(X)) \leq D_{Poin}\mathbb{E}|\nabla f(X)|^2 \tag{13}$$

for all locally Lipschitz bounded functions, then X satisfies (10) with $\gamma = 1$ and $L = C\sqrt{D_{Poin}}$ (recall that C always denotes a universal constant). Assume now that X satisfies the logarithmic Sobolev inequality

$$\mathrm{Ent} f^2(X) \leq D_{LS}\mathbb{E}|\nabla f(X)|^2 \tag{14}$$

for locally Lipschitz bounded functions, where for a nonnegative random variable Y,

$$\mathrm{Ent} Y = \mathbb{E}Y \log Y - \mathbb{E}Y \log(\mathbb{E}Y).$$

Then, by the results from [3], it follows that X satisfies (10) with $\gamma = 1/2$ and $L = \sqrt{D_{LS}/2}$.

We will now generalize this observation to measures satisfying the so-called modified logarithmic Sobolev inequality (introduced in [27]). We will present it in greater generality than needed for proving (10), since we will use it later (in Section 6) to prove refined concentration results for random vectors with independent Weibull coordinates.

Let $\beta \in (2, \infty)$. We will say that a random vector $Y \in \mathbb{R}^k$ satisfies a β-modified logarithmic Sobolev inequality if for every locally Lipschitz bounded positive function $f \colon \mathbb{R}^k \to \mathbb{R}$,

$$\mathrm{Ent} f^2(Y) \leq D_{LS_\beta}\Big(\mathbb{E}|\nabla f(Y)|^2 + \mathbb{E}\frac{|\nabla f(Y)|^\beta}{f(Y)^{\beta-2}}\Big). \tag{15}$$

Let us also introduce two quantities, measuring the length of the gradient in product spaces. Consider a locally Lipschitz function $f \colon \mathbb{R}^{mk} \to \mathbb{R}$, where we identify R^{mk} with the m-fold Cartesian product of $\mathbb{R}^k$. Let $x = (x_1, \ldots, x_m)$, where $x_i \in \mathbb{R}^k$. For each $i = 1, \ldots, m$, let $|\nabla_i f(x)|$ be the length of the gradient of f, treated as a function of x_i only, with the other coordinates fixed. Now for $r \geq 1$, set

$$|\nabla f(x)|_r = \Big(\sum_{i=1}^m |\nabla_i f(x)|^r\Big)^{1/r}.$$

Note that if f is differentiable at x, then $|\nabla f(x)|_2 = |\nabla f(x)|$ (the Euclidean length of the "true" gradient), whereas for $k = 1$ (and f differentiable), $|\nabla f(x)|_r$ is the ℓ_r^m norm of $\nabla f(x)$.

Theorem 3.4. *Let $\beta \in [2, \infty)$ and Y be a random vector in $\mathbb{R}^k$, satisfying (15). Consider a random vector $X = (X_1, \ldots, X_m)$ in $\mathbb{R}^{mk}$, where $X_1, \ldots, X_m$ are independent copies of Y. Then for any locally Lipschitz $f \colon \mathbb{R}^{mk} \to \mathbb{R}$ such that $f(X)$ is integrable, and $p \geq 2$,*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C_\beta D_{LS_\beta}^{1/2} p^{1/2}\Big\||\nabla f(X)|_2\Big\|_p + D_{LS_\beta}^{1/\beta} p^{1/\alpha}\Big\||\nabla f(X)|_\beta\Big\|_p, \tag{16}$$

where $\alpha = \frac{\beta}{\beta-1}$ is the Hölder conjugate of β.

In particular using the above theorem with $m = 1$ and $k = n$, we obtain the following

Corollary 3.5. *If X is a random vector in $\mathbb{R}^n$ which satisfies the β-modified log-Sobolev inequality (15), then it satisfies (10) with $\gamma = \frac{\beta-1}{\beta} \geq \frac{1}{2}$ and $L = C_\beta \max(D_{LS_\beta}^{1/2}, D_{LS_\beta}^{1/\beta})$.*

We remark that in the class of logarithmically concave random vectors, the β-modified log-Sobolev inequality is known to be equivalent to concentration for 1-Lipschitz functions of the form $\mathbb{P}(|f(X) - \mathbb{E}f(X)| \geq t) \leq 2\exp(-ct^{\beta/(\beta-1)})$ [51].

Proof of Theorem 3.4. By the tensorization property of entropy (see e.g. [45], Proposition 5.6) we get for all positive locally Lipschitz bounded functions $f\colon \mathbb{R}^{mk} \to \mathbb{R}$,

$$\mathrm{Ent}\, f^2(X) \leq D_{LS_\beta}\Big(\mathbb{E}|\nabla f(X)|_2^2 + \sum_{i=1}^{m} \mathbb{E}\frac{|\nabla_i f(X)|^\beta}{f(X)^{\beta-2}}\Big). \tag{17}$$

Following [3], consider now any locally Lipschitz bounded $f > 0$ and denote $F(t) = \mathbb{E}f(X)^t$. For $t > 2$,

$$F'(t) = \mathbb{E}\left(f(X)^t \log f(X)\right)$$

and

$$\frac{d}{dt}\left(\mathbb{E}f(X)^t\right)^{2/t} = \frac{d}{dt}F(t)^{2/t} = F(t)^{2/t} \cdot \frac{d}{dt}\left(\frac{2}{t}\log F(t)\right)$$

$$= F(t)^{2/t}\left(\frac{2}{t}\frac{F'(t)}{F(t)} - \frac{2}{t^2}\log F(t)\right) = \frac{2}{t^2}F(t)^{\frac{2}{t}-1}\left(tF'(t) - F(t)\log F(t)\right)$$

$$= \frac{2}{t^2}\left(\mathbb{E}f(X)^t\right)^{\frac{2}{t}-1}\left(\mathbb{E}\left(f(X)^t \log f(X)^t\right) - \left(\mathbb{E}f(X)^t\right)\log\left(\mathbb{E}f(X)^t\right)\right).$$

By (17) applied to the function $g = f^{t/2} = \varphi \circ f$ where $\varphi(u) = |u|^{t/2}$,

$$\frac{d}{dt}\left(\mathbb{E}f(X)^t\right)^{2/t} \leq \frac{2}{t^2}\left(\mathbb{E}f(X)^t\right)^{\frac{2}{t}-1} \cdot D_{LS_\beta}\Big(\mathbb{E}|\nabla(\varphi \circ f)(X)|_2^2 + \mathbb{E}|\nabla(\varphi \circ f)(X)|_\beta^\beta f(X)^{t(2-\beta)/2}\Big).$$

By the chain rule and the Hölder inequality for the pair of conjugate exponents $t/2, t/(t-2)$,

$$\mathbb{E}|\nabla(\varphi \circ f)(X)|_2^2 = \mathbb{E}\big(\,|\varphi'(f(X))| \cdot |\nabla f(X)|_2\,\big)^2$$

$$\leq \left(\mathbb{E}|\nabla f(X)|_2^t\right)^{2/t}\left(\mathbb{E}\left(\varphi'(f(X))\right)^{2t/(t-2)}\right)^{(t-2)/t}$$

$$= \left\||\nabla f(X)|_2\right\|_t^2 \cdot \left(\frac{t^2}{4}\right)\left(\mathbb{E}f(X)^t\right)^{1-\frac{2}{t}}.$$

Similarly, for $t \geq \beta$,

$$\mathbb{E}|\nabla(\varphi \circ f)(X)|_\beta^\beta f(X)^{t(2-\beta)/2} = \frac{t^\beta}{2^\beta}\mathbb{E}f(X)^{(t/2-1)\beta}|\nabla f(X)|_\beta^\beta f(X)^{t(2-\beta)/2}$$

$$= \frac{t^\beta}{2^\beta}\mathbb{E}f(X)^{t-\beta}|\nabla f(X)|_\beta^\beta$$

$$\leq \frac{t^\beta}{2^\beta}(\mathbb{E}f(X)^t)^{1-\beta/t}(\mathbb{E}|\nabla f(X)|_\beta^t)^{\beta/t}$$

$$= \frac{t^\beta}{2^\beta}(\mathbb{E}f(X)^t)^{1-\beta/t}\left\||\nabla f(X)|_\beta\right\|_t^\beta.$$

Thus we get for $\beta \leq t \leq p$,

$$\frac{d}{dt}\left(\mathbb{E}f(X)^t\right)^{2/t} \leq \frac{D_{LS_\beta}}{2}\left\||\nabla f(X)|_2\right\|_p^2 + \frac{D_{LS_\beta}}{2^{\beta-1}}t^{\beta-2}(\mathbb{E}f(X)^t)^{(2-\beta)/t}\left\||\nabla f(X)|_\beta\right\|_p^\beta.$$

Denote $a = \frac{D_{LS_\beta}}{2}\left\||\nabla f(X)|_2\right\|_p^2$, $b = \frac{D_{LS_\beta}}{2^{\beta-1}}\left\||\nabla f(X)|_\beta\right\|_p^\beta$, $g(t) = \left(\mathbb{E}f(X)^t\right)^{2/t}$. The above inequality can be written as

$$g^{\beta/2-1}\frac{d}{dt}g \leq g^{\beta/2-1}a + t^{\beta-2}b$$

11

for $t \in [\beta, p]$ or, denoting $G = g^{\beta/2}$,

$$\frac{d}{dt}G \le \frac{\beta}{2}(G^{(\beta-2)/\beta}a + t^{\beta-2}b).$$

For $\varepsilon > 0$ consider now the function $H_\varepsilon(t) = (g(\beta) + a(t - \beta) + b^{2/\beta}t^{2-2/\beta} + \varepsilon)^{\beta/2}$. We have

$$H_\varepsilon(\beta) > G(\beta)$$

and

$$\frac{d}{dt}H_\varepsilon(t) = \frac{\beta}{2}H_\varepsilon(t)^{(\beta-2)/\beta}(a + (2 - 2/\beta)t^{1-2/\beta}b^{2/\beta}) \ge \frac{\beta}{2}(H_\varepsilon(t)^{(\beta-2)/\beta}a + t^{\beta-2}b),$$

where we used the assumption $\beta \ge 2$. Using the last three inequalities together with the fact that for $t \ge 0$ the function $x \mapsto x^{(\beta-2)/2}a + t^{\beta-2}b$ is increasing on $[0, \infty)$ we obtain that $G(t) \le H_\varepsilon(t)$ for all $t \in [\beta, p]$, which by taking $\varepsilon \to 0^+$ implies that for $p \ge \beta$,

$$g(p) = G(p)^{2/\beta} \le H_0(p)^{2/\beta} \le g(\beta) + \frac{D_{LS_\beta}}{2}(p - \beta)\big\||\nabla f(X)|_2\big\|_p^2 + \frac{D_{LS_\beta}^{2/\beta}}{2}p^{2-2/\beta}\big\||\nabla f(X)|_\beta\big\|_p^2,$$

i.e.,

$$\|f(X)\|_p^2 \le \|f(X)\|_\beta^2 + \frac{D_{LS_\beta}}{2}(p - \beta)\big\||\nabla f(X)|_2\big\|_p^2 + \frac{D_{LS_\beta}^{2/\beta}}{2}p^{2-2/\beta}\big\||\nabla f(X)|_\beta\big\|_p^2. \qquad (18)$$

The above inequality has been proved so far for strictly positive, locally Lipschitz functions (the boundedness assumption can be easily removed by truncation and passage to the limit). For the case of a general locally Lipschitz function f, take any $\varepsilon > 0$ and consider $\tilde{f} = |f| + \varepsilon$. Since $\tilde{f}$ is strictly positive and locally Lipschitz, the above inequality holds also for $\tilde{f}$. Taking $\varepsilon \to 0^+$, we can now extend (18) to arbitrary locally Lipschitz f.

Finally, assume $f \colon \mathbb{R}^{mk} \to \mathbb{R}$ is locally Lipschitz and $f(X)$ is integrable. Applying (18) to $f - \mathbb{E}f(X)$ instead of f and taking the square root, we obtain

$$\|f(X) - \mathbb{E}f(X)\|_p \le \|f(X) - \mathbb{E}f(X)\|_\beta + \sqrt{D_{LS_\beta}(p - \beta)}\big\||\nabla f(X)|_2\big\|_p + D_{LS_\beta}^{1/\beta}p^{1/\alpha}\big\||\nabla f(X)|_\beta\big\|_p$$

for $p \ge \beta$. For $p \in [2, \beta]$, since (15) implies the Poincaré inequality with constant $D_{LS_\beta}/2$ (see Proposition 2.3. in [27]), we get

$$\|f(X) - \mathbb{E}f(X)\|_p \le CD_{LS_\beta}^{1/2}p\big\||\nabla f(X)|_2\big\|_p$$

(see the remark following (13)). These two estimates yield (16) with $C_\beta = C\sqrt{\beta}$. $\qquad \square$

3.1 Applications of Theorem 1.2

Let us now present certain applications of estimates established in the previous section. For simplicity we will restrict to the basic setting presented in Theorem 1.2.

3.1.1 Polynomials

A typical application of Theorem 1.2 would be to obtain tail inequalities for multivariate polynomials in the random vector X. The constants involved in such estimates do not depend on the dimension, but only on the degree of the polynomial. As already mentioned in the introduction, our results in this setting can be considered a transference of inequalities by Latała from the tetrahedral Gaussian case to the case of non-necessarily product random vectors and general polynomials.

3.1.2 Additive functionals and related statistics

We will now consider three related classes of additive statistics of a random vector, often arising in various problems.

Additive functionals   Let X be a random vector in $\mathbb{R}^n$ satisfying (3). For a function $f\colon \mathbb{R} \to \mathbb{R}$ define the random variable

$$Z_f = f(X_1) + \ldots + f(X_n). \tag{19}$$

It is classical and follows from (3) by a simple application of the Chebyshev inequality that if f is smooth with $\|f'\|_\infty \leq \alpha$, then for all $t > 0$,

$$\mathbb{P}\big(|Z_f - \mathbb{E}Z_f| \geq t\big) \leq e^2 \exp\Big(-\frac{t^2}{e^2 nL^2\alpha^2}\Big). \tag{20}$$

Using Theorem 1.2 we can easily obtain inequalities which hold if f is a polynomial-like function, i.e., if $\|f^{(D)}\|_\infty < \infty$ for some D. Note that the derivatives of the function $F(x_1,\ldots,x_n) = f(x_1) + \ldots + f(x_n)$ have a very simple diagonal form. In consequence, calculating their $\|\cdot\|_{\mathcal{J}}$ norms is simple. More precisely, we have

$$\mathbf{D}^d F(x) = \operatorname{diag}_d\Big(f^{(d)}(x_1), \ldots, f^{(d)}(x_n)\Big),$$

where $\operatorname{diag}_d(x_1,\ldots,x_n)$ stands for the d-indexed matrix $(a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ such that $a_{\mathbf{i}} = x_i$ if $i_1 = \ldots = i_d = i$ and 0 otherwise. It is easy to see that if $\mathcal{J} = \{[d]\}$, then $\|\operatorname{diag}_d(x_1,\ldots,x_n)\|_{\mathcal{J}} = \sqrt{x_1^2 + \ldots + x_n^2}$ and if $\#\mathcal{J} \geq 2$, then $\|\operatorname{diag}_d(x_1,\ldots,x_n)\|_{\mathcal{J}} = \max_{i\leq n}|x_i|$. Therefore we obtain the following corollary to Theorem 1.2. We will apply it in the next section to linear eigenvalue statistics of random matrices.

Corollary 3.6. *Let X be a random vector in $\mathbb{R}^n$ satisfying (3), $f\colon \mathbb{R} \to \mathbb{R}$ a $\mathcal{C}^D$ function, such that $\|f^{(D)}\|_\infty < \infty$ and Z_f is defined by (19). Then for all $t > 0$,*

$$\mathbb{P}(|Z_f - \mathbb{E}Z_f| \geq t) \leq 2\exp\Big(-\frac{1}{C_D}\min\Big(\frac{t^2}{L^{2D}n\|f^{(D)}\|_\infty^2}, \frac{t^{2/D}}{L^2\|f^{(D)}\|_\infty^{2/D}}\Big)\Big)$$

$$+ 2\exp\Big(-\frac{1}{C_D}\min_{1\leq d\leq D-1}\Big(\frac{t^2}{L^{2d}\sum_{i=1}^n(\mathbb{E}f^{(d)}(X_i))^2}\Big)\Big)$$

$$+ 2\exp\Big(-\frac{1}{C_D}\min_{2\leq d\leq D-1}\Big(\frac{t^{2/d}}{L^2\max_{i\leq n}|\mathbb{E}f^{(d)}(X_i)|^{2/d}}\Big)\Big).$$

Clearly the case $D = 1$ of the above corollary recovers up to constants (20). Moreover using the (yet unproven) Theorem 1.3 one can see that for $f(x) = x^D$ and X being a standard Gaussian vector in $\mathbb{R}^n$, the estimate of the corollary is optimal up to absolute constants (in this case, since Z_f is a sum of independent random variables, one can also use estimates from [32]).

Additive functionals of partial sums   Let us now consider a slightly more involved additive functional of the form

$$S_f = \sum_{i=1}^n f\Big(\sum_{j=1}^i X_j\Big). \tag{21}$$

Such random variables arise e.g., in the study of additive functionals of random walks (see e.g. [59, 16]). For simplicity we will only discuss what can be obtained directly for Lipschitz

functions f and what Theorem 1.2 gives for f with bounded second derivative. Let thus $F(x) = \sum_{i=1}^{n} f(\sum_{j=1}^{i} x_j)$. We have $\frac{\partial}{\partial x_i} F(x) = \sum_{l \geq i} f'(\sum_{j \leq l} x_j)$. Therefore

$$\big\| |\nabla F| \big\|_{\infty}^2 = \|f'\|_{\infty}^2 \sum_{i=1}^{n} (n - i + 1)^2 = \frac{1}{6} n(n+1)(2n+1)\|f'\|_{\infty}^2,$$

which, when combined with (3) and Chebyshev's inequality yields

$$\mathbb{P}(|S_f - \mathbb{E}S_f| \geq t) \leq 2\exp\Big(-\frac{t^2}{CL^2 n^3 \|f'\|_{\infty}^2}\Big).$$

Now, let us assume that $f \in \mathcal{C}^2$ and f'' is bounded. We have

$$|\mathbb{E}\nabla F(X)|^2 = \sum_{i=1}^{n} \Big(\sum_{l=i}^{n} \mathbb{E}f'\Big(\sum_{j=1}^{l} X_j\Big)\Big)^2$$

Moreover

$$\frac{\partial^2}{\partial x_i \partial x_j} F(x_1, \ldots, x_n) = \sum_{l \geq i \vee j} f''\Big(\sum_{k=1}^{l} x_k\Big)$$

and thus

$$\|\mathbf{D}^2 F(x)\|_{\{1,2\}}^2 = \sum_{i,j=1}^{n} \Big(\sum_{l=i \vee j}^{n} f''\Big(\sum_{k=1}^{l} x_k\Big)\Big)^2 \leq 2\|f''\|_{\infty}^2 \sum_{i=1}^{n}\sum_{j=i}^{n} (n-j+1)^2 \leq Cn^4 \|f''\|_{\infty}^2.$$

Since $\mathbf{D}^2 F$ is a symmetric bilinear form, we have

$$\|\mathbf{D}^2 F(x)\|_{\{1\}\{2\}} \leq \sup_{|\alpha| \leq 1} \sum_{i,j=1}^{n} \sum_{l=i \vee j}^{n} \Big|f''\Big(\sum_{k=1}^{l} x_k\Big)\Big| \alpha_i \alpha_j$$

$$\leq \sup_{|\alpha| \leq 1} \|f''\|_{\infty} \sum_{l=1}^{n} \Big(\sum_{i \leq l} \alpha_i\Big)^2 \leq \sup_{|\alpha| \leq 1} \|f''\|_{\infty} \sum_{l=1}^{n} l \sum_{i \leq l} \alpha_i^2 \leq Cn^2 \|f''\|_{\infty}.$$

Using the above estimates and Theorem 1.2 we obtain

$$\mathbb{P}(|S_f - \mathbb{E}S_f| \geq t) \leq 2\exp\Big(-\frac{1}{CL^2} \min\Big(\frac{t^2}{\sum_{i=1}^{n} \big(\sum_{l=i}^{n} \mathbb{E}f'(\sum_{j=1}^{l} X_j)\big)^2}, \frac{t}{n^2 \|f''\|_{\infty}}\Big)\Big).$$

To effectively bound the sub-Gaussian coefficient in the above inequality one should use some additional information about the structure of the vector X. For a given function f it is of order at most n^5, but if, e.g., the function f is even and X is symmetric, it clearly vanishes. In this case we get

$$\mathbb{P}(|S_f - \mathbb{E}S_f| \geq t) \leq 2\exp\Big(-\frac{1}{CL^2}\frac{t}{n^2 \|f''\|_{\infty}}\Big).$$

One can check that if for instance X is a standard Gaussian vector in $\mathbb{R}^n$ and $f(x) = x^2$ then this estimate is tight up to the value of the constant C.

U-statistics Our last application in this section will concern U-statistics (for simplicity of order 2) of the random vector X, i.e., random variables of the form

$$U = \sum_{i,j \leq n, i \neq j} h_{ij}(X_i, X_j),$$

14

where $h_{ij} \colon \mathbb{R}^2 \to \mathbb{R}$ are smooth functions. Without loss of generality let us assume that $h_{ij}(x, y) = h_{ji}(y, x)$.

A simple application of Chebyshev's inequality and (3) gives that if $\mathbf{D}h_{i,j}$ are uniformly bounded on $\mathbb{R}^2$ then for all $t > 0$,

$$\mathbb{P}(|U - \mathbb{E}U| \geq t) \leq 2 \exp \Big(-\frac{1}{CL^2} \frac{t^2}{\sum_{i=1}^{n} (\sum_{j \neq i} \frac{\partial}{\partial x} h_{ij}(x_i, x_j))^2} \Big)$$

$$\leq 2 \exp \Big(-\frac{1}{CL^2} \frac{t^2}{n^3 \max_{i \neq j} \|\frac{\partial}{\partial x} h_{ij}\|_\infty^2} \Big).$$

For h_{ij} of class $\mathcal{C}^2$ with bounded derivatives of second order, a direct application of Theorem 1.2 gives

$$\mathbb{P}(|U - \mathbb{E}U| \geq t) \leq 2 \exp \Big(-\frac{1}{C} \min \Big(\frac{t^2}{L^4 \alpha^2}, \frac{t^2}{L^2 \beta^2}, \frac{t}{L^2 \gamma} \Big) \Big),$$

where

$$\alpha^2 = \sup_{x \in \mathbb{R}^n} \Big\{ \sum_{i,j \leq n, i \neq j} \Big(\frac{\partial^2}{\partial x \partial y} h_{ij}(x_i, x_j) \Big)^2 + \sum_{i=1}^{n} \Big(\sum_{j \neq i} \frac{\partial^2}{\partial x^2} h_{ij}(x_i, x_j) \Big)^2 \Big\}$$

$$\leq n^2 \max_{i \neq j} \Big\| \frac{\partial^2}{\partial x \partial y} h_{ij} \Big\|_\infty + n^3 \max_{i \neq j} \Big\| \frac{\partial^2}{\partial x^2} h_{ij} \Big\|_\infty,$$

$$\beta^2 = \sum_{i=1}^{n} \Big(\sum_{j \neq i} \mathbb{E} \frac{\partial}{\partial x} h_{ij}(X_i, X_j) \Big)^2 \leq n^3 \max_{i \neq j} |\mathbb{E} \frac{\partial}{\partial x} h_{ij}(X_i, X_j)|^2,$$

$$\gamma = \sup_{x \in \mathbb{R}^n} \sup_{|\alpha|, |\beta| \leq 1} \Big\{ \sum_{i,j \leq n, i \neq j} \frac{\partial^2}{\partial x \partial y} h_{ij}(x_i, x_j) \alpha_i \beta_j + \sum_{i=1}^{n} \alpha_i \beta_i \sum_{j \neq i} \frac{\partial^2}{\partial x^2} h_{ij}(x_i, x_j) \Big\}$$

$$\leq n \Big(\max_{i \neq j} \Big\| \frac{\partial^2}{\partial x \partial y} h_{ij} \Big\|_\infty + \max_{i \neq j} \Big\| \frac{\partial^2}{\partial x^2} h_{ij} \Big\|_\infty \Big).$$

In particular, if $h_{ij} = h$, a function with bounded derivatives of second order, we get $\alpha^2 = \mathcal{O}(n^3)$, $\beta^2 = \mathcal{O}(n^3)$, $\gamma = \mathcal{O}(n)$, which shows that the oscillations of U are of order at most $\mathcal{O}(n^{3/2})$. In the case of U-statistics of independent random variables, generated by bounded h, this is a well known fact, corresponding to the CLT and classical Hoeffding inequalities for U-statistics. We remark that in the so called non-degenerate case, i.e. when $\mathrm{Var}(\mathbb{E}_X h(X, Y)) > 0$, $n^{3/2}$ is then indeed the right normalization in the CLT for U-statistics (see e.g. [23]).

3.1.3 Linear statistics of eigenvalues of random matrices

We will now use Corollary 3.6 to obtain tail inequalities for linear eigenvalue statistics of random Wigner matrices. We remark that one could also apply to the random matrix case the other inequalities considered in the previous section, obtaining in particular estimates on U-statistics of eigenvalues (which have been recently investigated by Lytova and Pastur [48]). We will focus on linear eigenvalues statistics (additive functionals in the language of the previous section) and obtain inequalities involving as a sub-Gaussian term a Sobolev norm of the function f with respect to the semicircle law (the limiting spectral distribution for Wigner ensembles). We refer the reader to the monographs [4, 6, 49, 54] for basic facts concerning random matrices.

Consider thus a real symmetric $n \times n$ random matrix A ($n \geq 2$) and let $\lambda_1 \leq \ldots \leq \lambda_n$ be its eigenvalues. We will be interested in concentration inequalities for functionals of the form

$$Z = \sum_{i=1}^{n} f(\lambda_i / \sqrt{n}).$$

In [30] Guionnet and Zeitouni obtained concentration inequalities for Z with Lipschitz f assuming that the entries of A are independent and satisfy the log-Sobolev inequality with some constant L. More specifically, they prove that for all $t > 0$,

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp\left(-\frac{t^2}{8L\|f'\|_\infty^2}\right).$$

(In fact they treat a more general case of banded matrices, but for simplicity we will focus on the basic case.)

As a corollary to Theorem 1.2 we present below an inequality which compliments the above result. Our aim is to replace the strong parameter $\|f'\|_\infty$ controlling the sub-Gaussian tail by a weaker Sobolev norm with respect to the semicircular law

$$d\rho(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}_{(-2,2)}(x)\,dx.$$

(recall that this is the limiting spectral distribution for Wigner matrices). Imposing additional smoothness assumptions on the function f it can be done in a window $|t| \leq c_f n$, where c_f depends on f.

Proposition 3.7. *Assume the entries of the matrix A are independent (modulo symmetry conditions), mean zero and variance one random variables, satisfying the logarithmic Sobolev inequality (14) with constant L^2. If f is $\mathcal{C}^2$ with bounded second derivative, then for all $t > 0$,*

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp\left(-\frac{1}{C_L}\left(\frac{t^2}{\int_{-2}^2 f'^2\,d\rho + n^{-2/3}\|f''\|_\infty^2} \wedge \frac{nt}{\|f''\|_\infty}\right)\right). \tag{22}$$

Remark The case $f(x) = x^2$ shows that under the assumptions of Proposition 3.7 one cannot expect a tail behaviour better than exponential for large t. Indeed, since $Z = \frac{1}{n}(\lambda_1^2 + \ldots + \lambda_n^2) = \frac{1}{n}\sum_{i,j\leq n} A_{ij}^2$, even if A is a matrix with standard Gaussian entries, then for all $t > 0$, $\mathbb{P}(|Z - \mathbb{E}Z| \geq t) > \frac{1}{C}\exp(-C(t^2 \wedge nt))$.

Remark A similar inequality to (22) holds in the case of Hermitian matrices with independent entries as well. In the proof given below one should invoke an appropriate result concerning the speed of convergence of the spectral distribution of Wigner matrices to the semicircular law.

Proof. Let us identify the random matrix A with a random vector $\tilde{A} = (A_{ij})_{1\leq i \leq j \leq n}$ having values in $\mathbb{R}^{n(n+1)/2}$ endowed with the standard Euclidean norm $|\tilde{A}| = \left(\sum_{1\leq i \leq j \leq n} A_{ij}^2\right)^{1/2}$. Note that $\|A\|_{\mathrm{HS}} \leq \sqrt{2}|\tilde{A}|$. By independence of coordinates of $\tilde{A}$ and the tensorization property of the logarithmic Sobolev inequality (see, e.g., [45, Corollary 5.7]), $\tilde{A}$ also satisfies (14) with constant L^2. Furthermore, by the Hoffman-Wielandt inequality (see, e.g., [4, Lemma 2.1.19]) which asserts that if B, C are two $n \times n$ real symmetric (or Hermitian) matrices and $\lambda_i(B), \lambda_i(C)$ resp. their eigenvalues arranged in nondecreasing order, then

$$\sum_{i=1}^n |\lambda_i(B) - \lambda_i(C)|^2 \leq \|B - C\|_{\mathrm{HS}}^2,$$

the map $\tilde{A} \mapsto (\lambda_1/\sqrt{n}, \ldots, \lambda_n/\sqrt{n}) \in \mathbb{R}^n$ is $\sqrt{2/n}$-Lipschitz. Therefore, the random vector $(\lambda_1/\sqrt{n}, \ldots, \lambda_n/\sqrt{n})$ satisfies (14) with constant $2L^2/n$. In consequence, by the results from [3] (see also Theorem 3.4), $(\lambda_1/\sqrt{n}, \ldots, \lambda_n/\sqrt{n})$ also satisfies (3) with constant $L/\sqrt{n}$. Applying Corollary 3.6 with $D = 2$ we obtain

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp\left(-\frac{1}{CL^2}\left(\frac{t^2}{n^{-1}\sum_{i=1}^n(\mathbb{E}f'(\lambda_i/\sqrt{n}))^2 + L^2 n^{-1}\|f''\|_\infty^2} \wedge \frac{nt}{\|f''\|_\infty}\right)\right). \tag{23}$$

In what follows we shall estimate from above the term $n^{-1}\sum_{i=1}^n(\mathbb{E}f'(\lambda_i/\sqrt{n}))^2$ from (23). First, by Jensen's inequality

$$\frac{1}{n}\sum_{i=1}^n(\mathbb{E}f'(\lambda_i/\sqrt{n}))^2 \le \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n f'(\lambda_i/\sqrt{n})^2\right) = \int_{\mathbb{R}}(f')^2 d\mu, \qquad (24)$$

where μ is the expected spectral measure of the matrix $n^{-1/2}A$. According to Wigner's theorem, for a fixed f, μ converges to the semicircular law as $n\to\infty$ and thus $\int_{\mathbb{R}}(f')^2\,d\mu \to \int_{-2}^2(f')^2\,d\rho$. A non-asymptotic bound on the term $\int_{\mathbb{R}}f'^2\,d\mu$ can be obtained using the result of Bobkov, Götze and Tikhomirov [12] on the speed of convergence of the expected spectral distribution of real Wigner matrices to the semicircular law. Since each entry of A satisfies the logarithmic Sobolev inequality with constant L^2, it also satisfies the Poincaré inequality with the same constant (see e.g. [45, Chapter 5]). Therefore Theorem 1.1 from [12] gives

$$\sup_{x\in\mathbb{R}}|F_\mu(x) - F_\rho(x)| \le C_L n^{-2/3}, \qquad (25)$$

where F_μ and F_ρ are the distribution functions of μ and ρ, respectively.

The decay of $1 - F_\mu(x)$ and $F_\mu(x)$ as $x\to\infty$ and $x\to-\infty$ (resp.) can be obtained using the sub-Gaussian concentration of $\lambda_n/\sqrt{n}$ and $\lambda_1/\sqrt{n}$, which is, e.g., a consequence of (3) for the vector of eigenvalues of $n^{-1/2}A$. For example, for any $t\ge 0$,

$$\mathbb{P}\left(\frac{\lambda_n}{\sqrt{n}} \ge \mathbb{E}\frac{\lambda_n}{\sqrt{n}} + t\right) \le 2\exp\left(-\frac{1}{C}\frac{nt^2}{L^2}\right). \qquad (26)$$

Using the classical technique of δ-nets for estimating the operator norm of a matrix (see e.g. [56]) and the fact that the entries of A are sub-Gaussian (as they satisfy the logarithmic Sobolev inequality) one gets $\mathbb{E}\lambda_n \le \mathbb{E}\|A\|_{\text{op}} \le CL\sqrt{n}$, which together with (26) yields

$$1 - F_\mu(CL + t) \le \mathbb{P}\left(\frac{\lambda_n}{\sqrt{n}} \ge CL + t\right) \le 2\exp\left(-\frac{1}{C}\frac{nt^2}{L^2}\right) \qquad (27)$$

for all $t\ge 0$. Clearly, the same inequality holds for $F(-CL - t)$. Integrating by parts,

$$\int_{\mathbb{R}} f'^2\,d\mu = \int_{\mathbb{R}} f'^2\,d\rho + \int_{\mathbb{R}}\left(f'(x)^2\right)'(F_\rho(x) - F_\mu(x))\,dx. \qquad (28)$$

Combining the uniform estimate (25) with (27) and using an elementary inequality $2xy \le x^2+y^2$, we estimate the last integral in (28) as follows:

$$\left|\int_{\mathbb{R}}\left(f'(x)^2\right)'(F_\mu(x) - F_\rho(x))\,dx\right|$$

$$\le \int_{\mathbb{R}}|2f'(x)f''(x)|\left(\|F_\mu - F_\rho\|_\infty \wedge 2\exp\left(-\frac{n}{C}\frac{\text{dist}(x,[-CL,CL])^2}{L^2}\right)\right)dx$$

$$\le \int_{\mathbb{R}} f'(x)^2\,d\nu(x) + \nu(\mathbb{R})\left\|f''\right\|_\infty^2, \quad (29)$$

where

$$d\nu(x) = C_L n^{-2/3} \wedge 2\exp\left(-\frac{\text{dist}(x,[-CL,CL])^2}{2\sigma^2}\right)dx, \qquad \text{and} \qquad \sigma^2 = \frac{CL^2}{2n}.$$

We proceed to estimate the two last terms from (29). Take $r > 0$ such that

$$2e^{-r^2/(2\sigma^2)} = C_L n^{-2/3} \qquad (30)$$

or put $r = 0$ if no such r exists. Note that if we assume $C_L \geq 1$, as we obviously can, then

$$r \leq CLn^{-1/2}\sqrt{\log n}. \tag{31}$$

We shall need the following estimates, which are easy consequences of the standard estimate for a Gaussian tail:

$$\int_r^\infty e^{-y^2/(2\sigma^2)}\,dy \leq C\sigma e^{-r^2/(2\sigma^2)} \leq C_L\sigma n^{-2/3} \leq C_L n^{-7/6}, \tag{32}$$

and

$$\int_r^\infty y^2 e^{-y^2/(2\sigma^2)}\,dy \leq \left(\int_0^\infty y^4 e^{-y^2/(2\sigma^2)}\,dy\right)^{1/2}\left(\int_r^\infty e^{-y^2/(2\sigma^2)}\,dy\right)^{1/2}$$
$$\leq C_L\sigma^{5/2}(\sigma n^{-2/3})^{1/2} \leq C_L n^{-11/6}. \tag{33}$$

Now, (30), (31) and (32) yield

$$\nu(\mathbb{R}) \leq (CL + r)C_L n^{-2/3} + 4\int_r^\infty e^{-y^2/(2\sigma^2)}\,dy \leq C_L n^{-2/3}. \tag{34}$$

We shall also need the estimate for $\int_{\mathbb{R}} x^2\,d\nu(x)$ which follows from (30), (31) and (33):

$$\int_{\mathbb{R}} x^2\,d\nu(x) = \frac{2}{3}(CL + r)^3 C_L n^{-2/3} + 4\int_r^\infty (CL + y)^2 e^{-y^2/(2\sigma^2)}\,dy \leq C_L n^{-2/3}. \tag{35}$$

In order to estimate $\int_{\mathbb{R}} f'^2\,d\nu$, take any $x_0 \in [-2, 2]$ such that $|f'(x_0)|^2 \leq \int_{-2}^2 f'^2\,d\rho$, and use $|f'(x)| \leq |f'(x_0)| + |x - x_0|\,\|f''\|_\infty$ to obtain

$$\int_{\mathbb{R}} f'(x)^2\,d\nu(x) \leq 2\Big(\int_{-2}^2 f'^2\,d\rho\Big)\nu(\mathbb{R}) + 2\,\|f''\|_\infty^2\int_{\mathbb{R}}|x - x_0|^2\,d\nu(x)$$
$$\leq 2\Big(\int_{-2}^2 f'^2\,d\rho\Big)\nu(\mathbb{R}) + 4\,\|f''\|_\infty^2 x_0^2\nu(\mathbb{R}) + 4\,\|f''\|_\infty^2\int_{\mathbb{R}} x^2\,d\nu(x).$$

Plugging (34) and (35) into the above yields

$$\int_{\mathbb{R}} f'(x)^2\,d\nu(x) \leq C_L n^{-2/3}\left(\int_{-2}^2 f'^2\,d\rho + \|f''\|_\infty^2\right). \tag{36}$$

In turn, plugging (34) and (36) into (29) and then combining with (28) we finally get

$$\int_{\mathbb{R}} f'^2\,d\mu \leq (1 + C_L n^{-2/3})\int_{\mathbb{R}} f'^2\,d\rho + C_L n^{-2/3}\,\|f''\|_\infty$$

which combined with (23) and (24) completes the proof. $\qquad\square$

Remark With some more work (using truncations or working directly on moments) one can extend the above proposition to the case, when $|f''(x)| \leq a(1 + |x|^k)$ for some non-negative integer k and $a \in \mathbb{R}$. In this case we obtain

$$\mathbb{P}\big(|Z - \mathbb{E}Z| \geq t\big) \leq 2\exp\left(-\left(\frac{t^2}{C_L\int_{-2}^2 f'^2\,d\rho + C_{L,k}n^{-2/3}a^2} \wedge \frac{n}{C_{L,k}}\left(\frac{t}{a}\right)^{\frac{2}{k+2}}\right)\right).$$

We also remark that to obtain the inequality (23) one does not have to use independence of the entries of A, it is enough to assume that the vector $\tilde{A}$ satisfies the inequality (3).

4 Two-sided estimates of moments for Gaussian polynomials

We will now prove Theorem 1.3, showing that in the case of general polynomials in Gaussian variables, the estimates of Theorem 1.2 are optimal (up to constants depending only on the degree of the polynomial). In the special case of tetrahedral polynomials this follows from Latała's Theorem 1.1 and the following result by Kwapień.

Theorem 4.1 (Kwapień, [39]). *If $X = (X_1, \ldots, X_n)$ where X_i are independent symmetric random variables, Q is a multivariate tetrahedral polynomial of degree D with coefficients in a Banach space E and Q_d is its homogeneous part of degree d, then for any symmetric convex function $\Phi \colon E \to \mathbb{R}_+$ and any $d \in \{0, 1, \ldots, D\}$,*

$$\mathbb{E}\Phi(Q_d(X)) \leq \mathbb{E}\Phi(C_d Q(X)).$$

Indeed, when combined with Theorem 1.1 and the triangle inequality, the above theorem gives the following

Corollary 4.2. *Let*

$$Z = \sum_{0 \leq d \leq D} \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}}^{(d)} g_{i_1} \cdots g_{i_d},$$

where $A_d = (a_{\mathbf{i}}^{(d)})_{\mathbf{i} \in [n]^d}$ is a d-indexed symmetric matrix of real numbers such that $a_{\mathbf{i}} = 0$ if $i_j = i_l$ for some $k \neq l$ (we adopt the convention that for $d = 0$ we have a single number $a_{\emptyset}^{(0)}$). Then for any $p \geq 2$,

$$C_D^{-1} \sum_{0 \leq d \leq D} \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A_d\|_{\mathcal{J}} \leq \|Z\|_p \leq C_D \sum_{0 \leq d \leq D} \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|A_d\|_{\mathcal{J}}.$$

The strategy of proof of Theorem 1.3 is very simple and relies on infinite divisibility of Gaussian random vectors, which will help us approximate the law of a general polynomial in Gaussian variables by the law of a tetrahedral polynomial, for which we will use Corollary 4.2.

It will be convenient to have the polynomial f represented as a combination of multivariate Hermite polynomials:

$$f(x_1, \ldots, x_n) = \sum_{d=0}^{D} \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} h_{d_1}(x_1) \cdots h_{d_n}(x_n), \tag{37}$$

where

$$\Delta_d^n = \{\mathbf{d} = (d_1, \ldots, d_n) \colon \forall_{k \in [n]} \, d_k \geq 0 \text{ and } d_1 + \cdots + d_n = d\}$$

and $h_d(x) = (-1)^d e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$ is the d-th Hermite polynomial.

Let $(W_t)_{t \in [0,1]}$ be a standard Brownian motion. Consider standard Gaussian random variables $g = W_1$ and, for any positive integer N,

$$g_{j,N} = \sqrt{N}(W_{\frac{j}{N}} - W_{\frac{j-1}{N}}), \quad j = 1, \ldots, N.$$

For any $d \geq 0$, we have the following representation of $h_d(g) = h_d(W_1)$ as a multiple stochastic integral (see [33, Example 7.12 and Theorem 3.21]),

$$h_d(g) = d! \int_0^1 \int_0^{t_d} \cdots \int_0^{t_2} dW_{t_1} \cdots dW_{t_{d-1}} dW_{t_d}.$$

Approximating the multiple stochastic integral leads to

$$h_d(g) = d! \lim_{N\to\infty} N^{-d/2} \sum_{1 \le j_1 < \cdots < j_d \le N} g_{j_1,N} \cdots g_{j_d,N}$$

$$= \lim_{N\to\infty} N^{-d/2} \sum_{\mathbf{j}\in[N]^{\underline{d}}} g_{j_1,N} \cdots g_{j_d,N}, \tag{38}$$

where the limit is in $L^2(\Omega)$ (see [33, Theorem 7.3. and formula (7.9)]) and actually the convergence holds in any L^p (see [33, Theorem 3.50]). We remark that instead of multiple stochastic integrals with respect to the Wiener process we could use the CLT for canonical U-statistics (see [23, Chapter 4.2]), however the stochastic integral framework seems more convenient as it allows to put all the auxiliary variables on the same probability space.

Now, consider n independent copies $(W_t^{(i)})_{t\in[0,1]}$ of the Brownian motion $(i = 1,\ldots,n)$ together with the corresponding Gaussian random variables: $g^{(i)} = W_1^{(i)}$ and, for $N \ge 1$,

$$g_{j,N}^{(i)} = \sqrt{N}(W_{\frac{j}{N}}^{(i)} - W_{\frac{j-1}{N}}^{(i)}), \quad j = 1,\ldots,N.$$

In the lemma below we state the representation of a multivariate Hermite polynomial in the variables $g^{(1)},\ldots,g^{(n)}$ as a limit of tetrahedral polynomials in the variables $g_{j,N}^{(i)}$. To this end introduce some more notation. Let

$$G^{(n,N)} = (g_{1,N}^{(1)},\ldots,g_{N,N}^{(1)}, \ g_{1,N}^{(2)},\ldots,g_{N,N}^{(2)}, \ \ldots, \ g_{1,N}^{(n)},\ldots,g_{N,N}^{(n)}) = (g_{j,N}^{(i)})_{(i,j)\in[n]\times[N]}$$

be a Gaussian vector with $n \times N$ coordinates. We identify here the set $[nN]$ with $[n] \times [N]$ via the bijection $(i,j) \leftrightarrow (i-1)N + j$. We will also identify the sets $([n]\times[N])^d$ and $[n]^d \times [N]^d$ in a natural way. For $d \ge 0$ and $\mathbf{d} \in \Delta_d^n$, let

$$I_{\mathbf{d}} = \big\{\mathbf{i} \in [n]^d \colon \forall_{l\in[n]} \#\mathbf{i}^{-1}(\{l\}) = d_l\big\},$$

and define a d-indexed matrix $B_{\mathbf{d}}^{(N)}$ of n^d blocks each of size N^d as follows: for $\mathbf{i} \in [n]^d$ and $\mathbf{j} \in [N]^d$,

$$\left(B_{\mathbf{d}}^{(N)}\right)_{(\mathbf{i},\mathbf{j})} = \begin{cases} \frac{d_1!\cdots d_n!}{d!} N^{-d/2} & \text{if } \mathbf{i} \in I_{\mathbf{d}} \text{ and } (\mathbf{i},\mathbf{j}) := \big((i_1,j_1),\ldots,(i_d,j_d)\big) \in ([n]\times[N])^{\underline{d}}, \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 4.3. *With the above notation, for any $p > 0$,*

$$\big\langle B_{\mathbf{d}}^{(N)}, (G^{(n,N)})^{\otimes d}\big\rangle \underset{N\to\infty}{\longrightarrow} h_{d_1}(g^{(1)})\cdots h_{d_n}(g^{(n)}) \quad \text{in } L^p(\Omega).$$

Proof. Using (38) for each $h_{d_i}(g^{(i)})$,

$$h_{d_1}(g^{(1)})\cdots h_{d_n}(g^{(n)})$$

$$= \lim_{N\to\infty} N^{-d/2} \sum_{\substack{(j_1^{(1)},\ldots,j_{d_1}^{(1)})\in[N]^{\underline{d_1}} \\ \vdots \\ (j_1^{(n)},\ldots,j_{d_n}^{(n)})\in[N]^{\underline{d_n}}}} \big(g_{j_1^{(1)},N}^{(1)} \cdots g_{j_{d_1}^{(1)},N}^{(1)}\big) \cdots \big(g_{j_1^{(n)},N}^{(n)} \cdots g_{j_{d_n}^{(n)},N}^{(n)}\big).$$

For each N, the right-hand side equals

$$\frac{1}{\#I_{\mathbf{d}}} N^{-d/2} \sum_{\mathbf{i}\in I_{\mathbf{d}}} \sum_{\substack{\mathbf{j}\in[N]^d \text{ s.t.} \\ (\mathbf{i},\mathbf{j})\in([n]\times[N])^{\underline{d}}}} g_{j_1,N}^{(i_1)} \cdots g_{j_d,N}^{(i_d)} = \big\langle B_{\mathbf{d}}^{(N)}, (G^{(n,N)})^{\otimes d}\big\rangle,$$

since $\#I_{\mathbf{d}} = \frac{d!}{d_1!\cdots d_n!}$. $\qquad\square$

Note that $B_{\mathbf{d}}^{(N)}$ is symmetric, i.e., for any $\mathbf{i} \in [n]^d$, $\mathbf{j} \in [N]^d$ if $\pi \colon [d] \to [d]$ is a permutation and $\mathbf{i}' \in [n]^d$, $\mathbf{j}' \in [N]^d$ are such that $\forall_{k \in [d]}\, i'_k = i_{\pi(k)}$ and $j'_k = j_{\pi(k)}$, then

$$\left(B_{\mathbf{d}}^{(N)}\right)_{(\mathbf{i}',\mathbf{j}')} = \left(B_{\mathbf{d}}^{(N)}\right)_{(\mathbf{i},\mathbf{j})}.$$

Moreover, $B_{\mathbf{d}}^{(N)}$ has zeros on "generalized diagonals", i.e., $\left(B_{\mathbf{d}}^{(N)}\right)_{(\mathbf{i},\mathbf{j})} = 0$ if $(i_k, j_k) = (i_l, j_l)$ for some $k \neq l$.

Proof of Theorem 1.3. Let us first note that it is enough to prove the moment estimates, the tail bound follows from them by the Paley-Zygmund inequality (see e.g. the proof of Corollary 1 in [43]). Moreover, the upper bound on moments follows directly from Theorem 1.2. For the lower bound we use Lemma 4.3 to approximate the L^p norm of $f(G) - \mathbb{E}f(G)$ with that of a tetrahedral polynomial, for which we can use the lower bound from Corollary 4.2.

Assuming f is of the form (37), Lemma 4.3 together with the triangle inequality implies

$$\lim_{N \to \infty} \left\| \sum_{d=1}^{D} \left\langle \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)}, \left(G^{(n,N)}\right)^{\otimes d} \right\rangle \right\|_p = \left\| f(G) - \mathbb{E}f(G) \right\|_p$$

for any $p > 0$, where $G = (g^{(1)}, \ldots, g^{(n)})$. It therefore remains to relate $\left\| \sum_{\mathbf{d}} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)} \right\|_{\mathcal{J}}$ with $\left\| \mathbb{E}\mathbf{D}^d f(G) \right\|_{\mathcal{J}}$ for any $d \geq 1$ and $\mathcal{J} \in P_d$. In fact we shall prove that

$$\lim_{N \to \infty} \left\| \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)} \right\|_{\mathcal{J}} = \frac{1}{d!} \left\| \mathbb{E}\mathbf{D}^d f(G) \right\|_{\mathcal{J}}, \tag{39}$$

which will end the proof.

Fix $d \geq 1$ and $\mathcal{J} \in P_d$. For any $\mathbf{d} \in \Delta_d^n$ define a symmetric d-indexed matrix $(b_{\mathbf{d}})_{\mathbf{i} \in [n]^d}$ as

$$(b_{\mathbf{d}})_{\mathbf{i}} = \begin{cases} \frac{d_1! \cdots d_n!}{d!} & \text{if } \mathbf{i} \in I_{\mathbf{d}}, \\ 0 & \text{otherwise.} \end{cases}$$

and a symmetric d-indexed matrix $(\tilde{B}_{\mathbf{d}}^{(N)})_{(\mathbf{i},\mathbf{j}) \in ([n] \times [N])^d}$ as

$$(\tilde{B}_{\mathbf{d}}^{(N)})_{(\mathbf{i},\mathbf{j})} = N^{-d/2} (b_{\mathbf{d}})_{\mathbf{i}} \quad \text{for all } \mathbf{i} \in [n]^d \text{ and } \mathbf{j} \in [N]^d.$$

It is a simple observation that

$$\left\| \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} \tilde{B}_{\mathbf{d}}^{(N)} \right\|_{\mathcal{J}} = \left\| \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} (b_{\mathbf{d}})_{\mathbf{i} \in [n]^d} \right\|_{\mathcal{J}}. \tag{40}$$

On the other hand, for any $\mathbf{d} \in \Delta_d^n$, the matrices $\tilde{B}_{\mathbf{d}}^{(N)}$ and $B_{\mathbf{d}}^{(N)}$ differ at no more than $\#I_{\mathbf{d}} \cdot \#([N]^d \setminus [N]^{\underline{d}})$ entries. More precisely, if $\mathcal{J}_0 = \{[d]\}$ (a trivial partition of $[d]$ into one set), then

$$\left\| \tilde{B}_{\mathbf{d}}^{(N)} - B_{\mathbf{d}}^{(N)} \right\|_{\mathcal{J}}^2 \leq \left\| \tilde{B}_{\mathbf{d}}^{(N)} - B_{\mathbf{d}}^{(N)} \right\|_{\mathcal{J}_0}^2 \leq \frac{d_1! \cdots d_n!}{d!} N^{-d}(N^d - N^{\underline{d}}) \longrightarrow 0 \quad \text{as } N \to \infty.$$

Thus the triangle inequality for the $\| \cdot \|_{\mathcal{J}}$ norm together with (40) yields

$$\lim_{N \to \infty} \left\| \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} B_{\mathbf{d}}^{(N)} \right\|_{\mathcal{J}} = \left\| \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} (b_{\mathbf{d}})_{\mathbf{i} \in [n]^d} \right\|_{\mathcal{J}}. \tag{41}$$

Finally, note that

$$\mathbb{E}\mathbf{D}^d f(G) = d! \sum_{\mathbf{d} \in \Delta_d^n} a_{\mathbf{d}} (b_{\mathbf{d}})_{\mathbf{i} \in [n]^d}. \tag{42}$$

21

Indeed, using the identity on Hermite polynomials, $h'_k(x) = kh_{k-1}(x)$ $(k \geq 1)$, we obtain $\mathbb{E}h_k^{(l)}(g) = k!\delta_{k,l}$ for $k, l \geq 0$, where $f^{(l)}$ stands for the l-th derivative of f, and thus, for any $\mathbf{d} \in \Delta_d^n$,

$$\left(\mathbb{E}\mathbf{D}^d h_{d_1}(g^{(1)}) \cdots h_{d_n}(g^{(n)})\right)_{\mathbf{i}} = d!(b_{\mathbf{d}})_{\mathbf{i}} \quad \text{for each } \mathbf{i} \in [n]^d.$$

Now, (42) follows by linearity. Combining it with (41) proves (39). $\qquad\square$

Remark Note that the above infinite-divisibility argument can be also used to prove the upper bound on moments in Theorem 1.3 (giving a proof independent of the one relying on Theorem 1.2).

5 Polynomials in independent sub-Gaussian random variables

In this section we prove Theorem 1.4. Before we proceed with the core of the proof we will need to introduce some auxiliary inequalities for the norms $\|\cdot\|_{\mathcal{J}}$ as well as some additional notation.

5.1 Properties of $\|\cdot\|_{\mathcal{J}}$ norms

The first inequality we will need is pretty standard and given in the following lemma (it is a direct consequence of the definition of the norms $\|\cdot\|_{\mathcal{J}}$).

Lemma 5.1. *For any d-indexed matrix $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ and any vectors $v_1, \ldots, v_d \in \mathbb{R}^n$ we have for all $\mathcal{J} \in P_d$,*

$$\|A \circ \otimes_{i=1}^d v_i\|_{\mathcal{J}} \leq \|A\|_{\mathcal{J}} \prod_{i=1}^d \|v_i\|_{\infty}$$

To formulate subsequent inequalities we need some auxiliary notation concerning d-indexed matrices. We will treat matrices as functions from $[n]^d$ into the real line, which in particular allows us to use the notation of indicator functions and for a set $C \subseteq \{1, \ldots, n\}^d$ write $\mathbf{1}_C$ for the matrix $(a_{\mathbf{i}})$ such that $a_{\mathbf{i}} = 1$ if $\mathbf{i} \in C$ and 0 otherwise.

Note that for $\#\mathcal{J} > 1$, $\|\cdot\|_{\mathcal{J}}$ is not unconditional in the standard basis, i.e., in general it is not true that $\|A \circ \mathbf{1}_C\|_{\mathcal{J}} \leq \|A\|_{\mathcal{J}}$. One situation in which this inequality holds is when C is of the form $C = \{\mathbf{i}: i_{k_1} = j_1, \ldots, i_{k_l} = j_l\}$ for some $1 \leq k_1 < \ldots < k_l \leq d$ and $j_1, \ldots, j_l \in [n]$ (which follows from Lemma 5.1). This corresponds to setting to zero all coefficients which are outside a "generalized row" of a matrix and leaving the coefficients in this row intact.

Later we will need another inequality of this type, which will allow us to select a "generalized diagonal" of a matrix. The corresponding estimate is given in the following

Lemma 5.2. *Let $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ be a d-indexed matrix and let $C \subseteq [n]^d$ be of the form $C = \{\mathbf{i}: i_k = i_l \text{ for } k, l \in K\}$, with $K \subseteq [d]$. Then for every $\mathcal{J} \in P_d$, $\|A \circ \mathbf{1}_C\|_{\mathcal{J}} \leq \|A\|_{\mathcal{J}}$.*

Proof. Since $\mathbf{1}_{C_1 \cap C_2} = \mathbf{1}_{C_1} \circ \mathbf{1}_{C_2}$, it is enough to consider the case $\#K = 2$, i.e. $C = \{\mathbf{i}: i_k = i_l\}$ for some $1 \leq k < l \leq d$. Let $\mathcal{J} = \{J_1, \ldots, J_m\}$. We will consider two cases.

1. The numbers k and l are separated by the partition $\mathcal{J}$. Without loss of generality we can assume that $k \in J_1$, $l \in J_2$. Then

$$\|A \circ \mathbf{1}_C\|_{\mathcal{J}} \tag{43}$$
$$= \sup_{\|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1: j \geq 3} \left(\sup_{\|x_{\mathbf{i}_{J_1}}^{(1)}\|_2, \|x_{\mathbf{i}_{J_2}}^{(2)}\|_2 \leq 1} \sum_{|\mathbf{i}_{J_1}| \leq n} \sum_{|\mathbf{i}_{J_2}| \leq n} \mathbf{1}_{\{i_k = i_l\}} \left(\sum_{|\mathbf{i}_{(J_1 \cup J_2)^c}| \leq n} a_{\mathbf{i}} x_{\mathbf{i}_{J_3}}^{(3)} \cdots x_{\mathbf{i}_{J_m}}^{(m)} \right) x_{\mathbf{i}_{J_1}}^{(1)} x_{\mathbf{i}_{J_2}}^{(2)} \right).$$

For any $x_{\mathbf{i}_{J_3}}^{(3)}, \ldots, x_{\mathbf{i}_{J_m}}^{(m)}$, consider the matrix

$$B_{\mathbf{i}_{J_1},\mathbf{i}_{J_2}} = \Big(\sum_{|\mathbf{i}_{(J_1 \cup J_2)^c}| \leq n} a_{\mathbf{i}} x_{\mathbf{i}_{J_3}}^{(3)} \cdots x_{\mathbf{i}_{J_m}}^{(m)} \Big)_{\mathbf{i}_{J_1},\mathbf{i}_{J_2}}$$

acting from $\ell_2([n]^{J_1})$ to $\ell_2([n]^{J_2})$.

For fixed $x_{\mathbf{i}_{J_3}}^{(3)}, \ldots, x_{\mathbf{i}_{J_m}}^{(m)}$ the inner expression on the right hand side of (43) is the operator norm of the block-diagonal matrix obtained from $B_{\mathbf{i}_{J_1},\mathbf{i}_{J_2}}$ by setting to zero entries in off-diagonal blocks. Therefore it is not greater than the operator norm of $B_{\mathbf{i}_{J_1},\mathbf{i}_{J_2}}$, which allows us to write

$$\|A \circ \mathbf{1}_C\|_{\mathcal{J}} \leq \sup_{\|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1: \, j \geq 3} \Big(\sup_{\|x_{\mathbf{i}_{J_1}}^{(1)}\|_2, \|x_{\mathbf{i}_{J_2}}^{(2)}\|_2 \leq 1} \sum_{|\mathbf{i}_{J_1}| \leq n} \sum_{|\mathbf{i}_{J_2}| \leq n} \Big(\sum_{|\mathbf{i}_{(J_1 \cup J_2)^c}| \leq n} a_{\mathbf{i}} x_{\mathbf{i}_{J_3}}^{(3)} \cdots x_{\mathbf{i}_{J_m}}^{(m)} \Big) x_{\mathbf{i}_{J_1}}^{(1)} x_{\mathbf{i}_{J_2}}^{(2)} \Big)$$

$$= \|A\|_{\mathcal{J}}.$$

2. There exists j such that $k, l \in J_j$. Without loss of generality we can assume that $j = 1$. We have

$$\|A \circ \mathbf{1}_C\|_{\mathcal{J}} = \sup_{\|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1: \, j \geq 2} \Big(\sup_{\|x_{\mathbf{i}_{J_1}}^{(1)}\|_2 \leq 1} \sum_{|\mathbf{i}_{J_1}| \leq n} \mathbf{1}_{\{i_k = i_l\}} \Big(\sum_{|\mathbf{i}_{J_1^c}| \leq n} a_{\mathbf{i}} x_{\mathbf{i}_{J_2}}^{(2)} \cdots x_{\mathbf{i}_{J_m}}^{(m)} \Big) x_{\mathbf{i}_{J_1}}^{(1)} \Big)$$

$$= \sup_{\|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1: \, j \geq 2} \Big(\sum_{|\mathbf{i}_{J_1}| \leq n} \mathbf{1}_{\{i_k = i_l\}} \Big(\sum_{|\mathbf{i}_{J_1^c}| \leq n} a_{\mathbf{i}} x_{\mathbf{i}_{J_2}}^{(2)} \cdots x_{\mathbf{i}_{J_m}}^{(m)} \Big)^2 \Big)^{1/2}$$

$$\leq \sup_{\|x_{\mathbf{i}_{J_j}}^{(j)}\|_2 \leq 1: \, j \geq 2} \Big(\sum_{|\mathbf{i}_{J_1}| \leq n} \Big(\sum_{|\mathbf{i}_{J_1^c}| \leq n} a_{\mathbf{i}} x_{\mathbf{i}_{J_2}}^{(2)} \cdots x_{\mathbf{i}_{J_m}}^{(m)} \Big)^2 \Big)^{1/2} = \|A\|_{\mathcal{J}}.$$

$\square$

For a partition $\mathcal{K} = \{K_1, \ldots, K_m\} \in P_d$ define

$$L(\mathcal{K}) = \{\mathbf{i} \in [n]^d : i_k = i_l \text{ iff } \exists_{j \leq m} \, k, l \in K_j\}. \tag{44}$$

Thus $L(\mathcal{K})$ is the set of all indices for which the partition into level sets is equal to $\mathcal{K}$.

Corollary 5.3. *For any $\mathcal{J}, \mathcal{K} \in P_d$ and any d-indexed matrix A,*

$$\|A \circ \mathbf{1}_{L(\mathcal{K})}\|_{\mathcal{J}} \leq 2^{\#\mathcal{K}(\#\mathcal{K}-1)/2} \|A\|_{\mathcal{J}}.$$

Proof. By Lemma 5.2 and the triangle inequality for any $k < l$, $\|A \circ \mathbf{1}_{\{i_k \neq i_l\}}\|_{\mathcal{J}} = \|A - A \circ \mathbf{1}_{\{i_k = i_l\}}\|_{\mathcal{J}} \leq 2\|A\|_{\mathcal{J}}$. Now it is enough to note that $L(\mathcal{K})$ can be expressed as an intersection of $\#\mathcal{K}$ "generalized diagonals" and $\#\mathcal{K}(\#\mathcal{K}-1)/2$ sets of the form $\{\mathbf{i} : i_k \neq i_l\}$ where $k < l$ and use again Lemma 5.2 together with the above inequality. $\square$

5.2 Proof of Theorem 1.4

Let us first note that the tail bound of Theorem 1.4 follows from the moment estimate and Chebyshev inequality in the same way as in Theorems 1.2 or 3.3. We will therefore focus on the moment bound.

The method of proof will rely on the reduction to the Gaussian case via decoupling inequalities, symmetrization and the contraction principle. To carry out this strategy we will need the following representation of f.

$$f(x) = \sum_{0 \leq d \leq D} \sum_{m=0}^{d} \sum_{\substack{k_1, \ldots, k_m > 0 \\ k_1 + \ldots + k_m = d}} \sum_{\mathbf{i} \in [n]^{\underline{m}}} c_{(i_1, k_1), \ldots, (i_m, k_m)}^{(d)} x_{i_1}^{k_1} x_{i_2}^{k_2} \cdots x_{i_m}^{k_m}, \tag{45}$$

23

where the coefficients $c^{(d)}_{(i_1,k_1),\dots,(i_m,k_m)}$ satisfy

$$c^{(d)}_{(i_1,k_1),\dots,(i_m,k_m)} = c^{(d)}_{(i_{\pi_1},k_{\pi_1}),\dots,(i_{\pi_m},k_{\pi_m})} \tag{46}$$

for all permutations $\pi\colon [m] \to [m]$. At this point we would like to explain the convention regarding indices which we will use throughout this section. It is rather standard, but we prefer to draw the Reader's attention to it, as we will use it extensively in what follows. Namely, we will treat the sequence $\mathbf{k} = (k_1,\dots,k_m)$ as a function acting on $[m]$ and taking values in positive integers. In particular if $m = 0$, then $[m] = \emptyset$ and there exists exactly one function $\mathbf{k}\colon [m] \to \mathbb{N} \setminus \{0\}$ (the empty function). Moreover by convention this function satisfies $\sum_{i=1}^{m} k_i = 0$ (as the summation runs over an empty set). Therefore, for $d = 0$ and $m = 0$ the subsum over $k_1,\dots,k_m$ and $\mathbf{i}$ above is equal to the free coefficient of the polynomial (which can be denoted by $c^{(0)}_{\emptyset}$), since the summation over $k_1,\dots,k_m$ runs over a one-element set containing the empty index/function and for this index there is exactly one index $\mathbf{i}\colon [m] \to \{1,\dots,n\}$, which belongs to $[n]^{\underline{m}}$ (again the empty-index). Here we also use the convention that a product over an empty set is equal to one. On the other hand, for $d > 0$, the contribution from $m = 0$ is equal to zero (as the empty index $\mathbf{k}$ does not satisfy the constraint $k_1 + \dots + k_m = d$ and so the summation over $k_1,\dots,k_m$ runs over the empty set).

Using (45) together with independence of $X_1,\dots,X_n$, one may write

$$f(X) - \mathbb{E}f(X) = \sum_{1 \le d \le D} \sum_{m=1}^{d} \sum_{\substack{k_1,\dots,k_m>0 \\ k_1+\dots+k_m=d}} \sum_{\mathbf{i}\in[n]^{\underline{m}}} c^{(d)}_{(i_1,k_1),\dots,(i_m,k_m)} \sum_{\emptyset \ne J \subseteq [m]} \prod_{j\in J}(X_{i_j}^{k_j} - \mathbb{E}X_{i_j}^{k_j}) \prod_{j\notin J}\mathbb{E}X_{i_j}^{k_j}.$$

Rearranging the terms and using (46) together with the triangle inequality, we obtain

$$|f(X) - \mathbb{E}f(X)| \le \sum_{1 \le d \le D} \sum_{a=1}^{d} \sum_{\substack{k_1,\dots,k_a>0 \\ k_1+\dots+k_a=d}} \Big| \sum_{\mathbf{i}\in[n]^{\underline{a}}} d^{(k_1,\dots,k_a)}_{i_1,\dots,i_a}(X_{i_1}^{k_1} - \mathbb{E}X_{i_1}^{k_1})\cdots(X_{i_a}^{k_a} - \mathbb{E}X_{i_a}^{k_a})\Big|,$$

where

$$d^{(k_1,\dots,k_a)}_{i_1,\dots,i_a} = \sum_{a\le m\le D} \sum_{\substack{k_{a+1},\dots,k_m>0: \\ k_1+\dots+k_m\le D}} \sum_{\substack{i_{a+1},\dots,i_m: \\ (i_1,\dots,i_m)\in[n]^{\underline{m}}}} \binom{m}{a} c^{(k_1+\dots+k_m)}_{(i_1,k_1),\dots,(i_m,k_m)} \mathbb{E}X_{i_{a+1}}^{k_{a+1}}\cdots\mathbb{E}X_{i_m}^{k_{i_m}}.$$

Note that (46) implies that for every permutation $\pi\colon [a] \to [a]$,

$$d^{(k_1,\dots,k_a)}_{i_1,\dots,i_a} = d^{(k_{\pi_1},\dots,k_{\pi_a})}_{i_{\pi_1},\dots,i_{\pi_a}}. \tag{47}$$

Let now $X^{(1)},\dots,X^{(D)}$ be independent copies of the random vector X and $(\varepsilon_i^{(j)})_{i\le n,j\le D}$ an array of i.i.d. Rademacher variables independent of $(X^{(j)})_j$. For each $k_1,\dots,k_a$, by decoupling inequalities (Theorem 7.1 in the Appendix) applied to the functions

$$h^{(k_1,\dots,k_a)}_{i_1,\dots,i_a}(x_1,\dots,x_a) = d^{(k_1,\dots,k_a)}_{i_1,\dots,i_a}(x_1^{k_1} - \mathbb{E}X_{i_1}^{k_1})\cdots(x_a^{k_a} - \mathbb{E}X_{i_a}^{k_a})$$

and standard symmetrization inequalities (applied conditionally a times) we obtain,

$$\|f(X) - \mathbb{E}f(X)\|_p \tag{48}$$

$$\le C_D \sum_{d=1}^{D}\sum_{a=1}^{d} \sum_{\substack{k_1,\dots,k_a>0 \\ k_1+\dots+k_a=d}} \Big\| \sum_{\mathbf{i}\in[n]^{\underline{a}}} d^{(k_1,\dots,k_a)}_{i_1,\dots,i_a}\Big((X_{i_1}^{(1)})^{k_1} - \mathbb{E}(X_{i_1}^{(1)})^{k_1}\Big)\cdots\Big((X_{i_a}^{(a)})^{k_a} - \mathbb{E}(X_{i_a}^{(a)})^{k_a}\Big)\Big\|_p$$

$$\le C_D \sum_{d=1}^{D}\sum_{a=1}^{d} \sum_{\substack{k_1,\dots,k_a>0 \\ k_1+\dots+k_a=d}} \Big\| \sum_{\mathbf{i}\in[n]^{\underline{a}}} d^{(k_1,\dots,k_a)}_{i_1,\dots,i_a}\Big(\varepsilon_{i_1}^{(1)}(X_{i_1}^{(1)})^{k_1}\cdots\varepsilon_{i_a}^{(a)}(X_{i_a}^{(a)})^{k_a}\Big)\Big\|_p$$

(note that in the first part of Theorem 7.1 one does not impose any symmetry assumptions on the functions $h_{\mathbf{i}}$).

We will now use the following standard comparison lemma (for reader's convenience its proof is presented in the Appendix).

Lemma 5.4. *For any positive integer k, if $Y_1, \ldots, Y_n$ are independent symmetric variables with $\|Y_i\|_{\psi_{2/k}} \leq M$, then*

$$\| \sum_{i=1}^n a_i Y_i \|_p \leq C_k M \| \sum_{i=1}^n a_i g_{i1} \cdots g_{ik} \|_p,$$

where g_{ij} are i.i.d. $\mathcal{N}(0,1)$ variables.

Note that for any positive integer k we have $\|X_i^k\|_{\psi_{2/k}} = \|X_i\|_{\psi_2}^k \leq L^k$, so (48) together with the above lemma (used repeatedly and conditionally) yield

$$\|f(X) - \mathbb{E}f(X)\|_p$$

$$\leq C_D \sum_{1 \leq d \leq D} L^d \sum_{a=1}^d \sum_{\substack{k_1, \ldots, k_a > 0 \\ k_1 + \ldots + k_a = d}} \Big\| \sum_{\mathbf{i} \in [n]^{\underline{a}}} d_{i_1, \ldots, i_a}^{(k_1, \ldots, k_a)} (g_{i_1,1}^{(1)} \cdots g_{i_1,k_1}^{(1)}) \cdots (g_{i_a,1}^{(a)} \cdots g_{i_a,k_a}^{(a)}) \Big\|_p, \quad (49)$$

where $(g_{i,k}^{(j)})$ is an array of i.i.d. standard Gaussian variables. Consider now multi-indexed matrices $B_1, \ldots, B_D$ defined as follows. For $1 \leq d \leq D$, and a multi-index $\mathbf{r} = (r_1, \ldots, r_d) \in [n]^d$ let $\mathcal{I} = \{I_1, \ldots, I_a\}$ be the partition of $\{1, \ldots, d\}$ into the level sets of $\mathbf{r}$ and $i_1, \ldots, i_a$ be the values corresponding to the level sets $I_1, \ldots, I_a$. Define moreover

$$b_{r_1, \ldots, r_d}^{(d)} = d_{i_1, \ldots, i_a}^{(\#I_1, \ldots, \#I_a)}$$

(note that thanks to (47) this definition does not depend on the order of $I_1, \ldots, I_a$). Finally, define the d-indexed matrix $B_d = (b_{\mathbf{r}}^{(d)})_{\mathbf{r} \in [n]^d}$.

Let us also define for $k_1, \ldots, k_a > 0$, $\sum_{i=1}^a k_i = d$ the partition $\mathcal{K}(k_1, \ldots, k_a) \in \mathcal{P}_d$ by splitting the set $\{1, \ldots, d\}$ into consecutive intervals of length $k_1, \ldots, k_a$, i.e., $\mathcal{K} = \{K_1, \ldots, K_a\}$, where for $l = 1, \ldots, a$, $K_l = \{1 + \sum_{i=1}^{l-1} k_i, 2 + \sum_{i=1}^{l-1} k_i, \ldots, \sum_{i=1}^l k_i\}$.

Applying Theorem 3.1 to the right hand side of (49), we obtain

$$\|f(X) - \mathbb{E}f(X)\|_p$$

$$\leq C_D \sum_{1 \leq d \leq D} L^d \sum_{a=1}^d \sum_{\substack{k_1, \ldots, k_a > 0 \\ k_1 + \ldots + k_a = d}} \Big\| \Big\langle B_d \circ \mathbf{1}_{L(\mathcal{K}(k_1, \ldots, k_a))}, \bigotimes_{j=1}^a \bigotimes_{k=1}^{k_j} (g_{i,k_j}^{(j)})_{i \leq n} \Big\rangle \Big\|_p$$

$$\leq C_D \sum_{1 \leq d \leq D} L^d \sum_{a=1}^d \sum_{\substack{k_1, \ldots, k_a > 0 \\ k_1 + \ldots + k_a = d}} \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|B_d \circ \mathbf{1}_{L(\mathcal{K}(k_1, \ldots, k_a))}\|_{\mathcal{J}}.$$

Note that for all $k_1, \ldots, k_a$ by Corollary 5.3 we have $\|B_d \circ \mathbf{1}_{L(\mathcal{K}(k_1, \ldots, k_a))}\|_{\mathcal{J}} \leq C_d \|B_d\|_{\mathcal{J}}$. Thus we obtain

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C_D \sum_{1 \leq d \leq D} L^d \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|B_d\|_{\mathcal{J}}.$$

Our next goal is to replace B_d in the above inequality by $\mathbb{E}\mathbf{D}^d f(X)$. To this end we will analyse the structure of the coefficients of B_d and compare them with the integrated partial derivatives of f.

Let us first calculate $\mathbb{E}\mathbf{D}^d f(X)$. Consider $\mathbf{r} \in [n]^d$, such that $i_1, \ldots, i_a$ are its distinct values, taken $l_1, \ldots, l_a$ times respectively. We have

$$
\mathbb{E}\frac{\partial^d f}{\partial x_{r_1} \cdots \partial x_{r_d}}(X) = \sum_{k_1 \geq l_1, \ldots, k_a \geq l_a} \sum_{a \leq m \leq D} \sum_{\substack{k_{a+1}, \ldots, k_m > 0 \\ k_1 + \ldots + k_m \leq D}} \sum_{\substack{i_{a+1}, \ldots, i_m \\ (i_1, \ldots, i_m) \in [n]^{\underline{m}}}}
$$
$$
\left[\binom{m}{a} a! c^{(k_1 + \ldots + k_m)}_{(i_1, k_1), \ldots, (i_m, k_m)} \prod_{j=1}^{a} \mathbb{E}X_{i_j}^{k_j - l_j} \prod_{j=a+1}^{m} \mathbb{E}X_{i_j}^{k_j} \prod_{j=1}^{a} \frac{k_j!}{(k_j - l_j)!} \right],
$$

where we have used (46).

By comparing this with the definition of $b^{(d)}_{r_1, \ldots, r_d}$ and $d^{(k_1, \ldots, k_a)}_{i_1, \ldots, i_a}$ one can see that the sub-sum of the right hand side above corresponding to the choice $k_1 = l_1, \ldots, k_a = l_a$ is equal to $a! l_1! \cdots l_a! b^{(d)}_{r_1, \ldots, r_d}$.

In particular for $d = D$, since $l_1 + \ldots + l_a = D$, we have

$$
\mathbb{E}\frac{\partial^D f}{\partial x_{r_1} \cdots \partial x_{r_D}}(X) = a! l_1! \cdots l_a! b^{(D)}_{r_1, \ldots, r_D}
$$

and so

$$
\|B_D\|_{\mathcal{J}} \leq \sum_{\mathcal{K} \in \mathcal{P}_D} \|B_D \circ \mathbf{1}_{L(\mathcal{K})}\|_{\mathcal{J}} \leq \sum_{\mathcal{K} \in \mathcal{P}_D} \|\mathbf{D}^D f(X) \circ \mathbf{1}_{L(\mathcal{K})}\|_{\mathcal{J}} \leq C_D \|\mathbf{D}^D f(X)\|_{\mathcal{J}},
$$

where in the last inequality we used Corollary 5.3. Therefore if we prove that for all $d < D$ and all partitions $\mathcal{I} = \{I_1, \ldots, I_a\}, \mathcal{J} = \{J_1, \ldots, J_b\} \in P_d$,

$$
\|a! \#I_1! \cdots \#I_a! (B_d \circ \mathbf{1}_{L(\mathcal{I})}) - \mathbb{E}\mathbf{D}^d f(X) \circ \mathbf{1}_{L(\mathcal{I})}\|_{\mathcal{J}} \leq C_D \sum_{d < k \leq D} L^{k-d} \sum_{\substack{\mathcal{K} \in P_k \\ \#\mathcal{K} = \#\mathcal{J}}} \|B_k\|_{\mathcal{K}}, \qquad (50)
$$

then by simple reverse induction (using again Corollary 5.3) we will obtain

$$
\sum_{1 \leq d \leq D} L^d \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|B_d\|_{\mathcal{J}} \leq C_D \sum_{1 \leq d \leq D} L^d \sum_{\mathcal{J} \in P_d} p^{\#\mathcal{J}/2} \|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}},
$$

which will end the proof of the theorem.

Fix any $d < D$ and partitions $\mathcal{I} = \{I_1, \ldots, I_a\}, \mathcal{J} = \{J_1, \ldots, J_b\} \in P_d$. Denote $l_i = \#I_i$. For every sequence $k_1, \ldots, k_a$ such that $k_i \geq l_i$ for $i \leq a$ and there exists $i \leq a$ such that $k_i > l_i$, let us define a d-indexed matrix $E^{(d, k_1, \ldots, k_a)}_{\mathcal{I}} = (e^{(d, k_1, \ldots, k_a)}_{\mathbf{r}})_{\mathbf{r} \in [n]^d}$, such that $e^{(d, k_1, \ldots, k_a)}_{\mathbf{r}} = 0$ if $\mathbf{r} \notin L(\mathcal{I})$ and for $\mathbf{r} \in L(\mathcal{I})$,

$$
e^{(d, k_1, \ldots, k_a)}_{\mathbf{r}} = \sum_{a \leq m \leq D} \sum_{\substack{k_{a+1}, \ldots, k_m > 0 \\ k_1 + \ldots + k_m \leq D}} \sum_{\substack{i_{a+1}, \ldots, i_m \\ (i_1, \ldots, i_m) \in [n]^{\underline{m}}}} \binom{m}{a} c^{(k_1 + \ldots + k_m)}_{(i_1, k_1), \ldots, (i_m, k_m)} \prod_{j=1}^{a} \mathbb{E}X_{i_j}^{k_j - l_j} \prod_{j=a+1}^{m} \mathbb{E}X_{i_j}^{k_j},
$$

where $i_1, \ldots, i_a$ are the values of $\mathbf{r}$ corresponding to the level sets $I_1, \ldots, I_a$. We then have

$$
\sum_{\substack{k_1 \geq l_1, \ldots, k_a \geq l_a \\ \exists_i k_i > l_i}} a! \frac{k_1!}{(k_1 - l_1)!} \cdots \frac{k_a!}{(k_a - l_a)!} E^{(d, k_1, \ldots, k_a)}_{\mathcal{I}} = \mathbb{E}\mathbf{D}^d f(X) \circ \mathbf{1}_{L(\mathcal{I})} - a! l_1! \cdots l_a! B_d \circ \mathbf{1}_{L(\mathcal{I})}.
$$

Since we do not pay attention to constants depending only on D, by the above formula and the triangle inequality, to prove (50) it is enough to show that for all sequences $k_1, \ldots, k_a$ such that $k_i \geq l_i$ for $i \leq a$ and there exists $i \leq a$ such that $k_i > l_i$ one has

$$
\|E^{(d, k_1, \ldots, k_a)}_{\mathcal{I}}\|_{\mathcal{J}} \leq C_D L^{\sum_{j \leq a}(k_j - l_j)} \|B_{k_1 + \ldots + k_a}\|_{\mathcal{K}} \qquad (51)
$$

for some partition $\mathcal{K} \in \mathcal{P}_{k_1+\ldots+k_a}$ with $\#\mathcal{K} = \#\mathcal{J}$ (note that $\sum_{j \le a} l_j = d$). Therefore in what follows we will fix $k_1, \ldots, k_a$ as above and to simplify the notation we will write $E^{(d)}$ instead of $E_{\mathcal{I}}^{(d,k_1,\ldots,k_a)}$ and $e_{\mathbf{r}}^{(d)}$ instead of $e_{\mathbf{r}}^{(d,k_1,\ldots,k_a)}$.

Fix therefore any partition $\tilde{\mathcal{I}} = \{\tilde{I}_1, \ldots, \tilde{I}_a\} \in \mathcal{P}_{k_1+\ldots+k_a}$ such that $\#\tilde{I}_i = k_i$ and $I_i \subseteq \tilde{I}_i$ for all $i \le a$ (the specific choice of $\tilde{\mathcal{I}}$ is irrelevant). Finally define a $(k_1 + \ldots + k_a)$-indexed matrix $\tilde{E}^{(k_1+\ldots+k_a)} = (\tilde{e}_{\mathbf{r}}^{(k_1+\ldots+k_a)})_{\mathbf{r} \in [n]^d}$ by setting

$$\tilde{e}_{\mathbf{r}}^{(k_1+\ldots+k_a)} = e_{\mathbf{r}_{[d]}}^{(d)} \mathbf{1}_{\{\mathbf{r} \in L(\tilde{\mathcal{I}})\}}. \tag{52}$$

In other words the new matrix is created by embedding the d-indexed matrix into a "generalized diagonal" of a $(k_1 + \ldots + k_a)$-indexed matrix by adding $\sum_{j \le a}(k_j - l_j)$ new indices and assigning to them the values of old indices (for each $j \le a$ we add $k_j - l_j$ times the common value attained by $\mathbf{r}_{\{1,\ldots,d\}}$ on I_j).

Recall now the definition of the coefficients $b_{\mathbf{r}}^{(d)}$ and note that for any $\mathbf{r} \in L(\tilde{\mathcal{I}}) \subseteq [n]^{k_1+\ldots+k_a}$ we have $\tilde{e}_{\mathbf{r}}^{(k_1+\ldots+k_a)} = b_{\mathbf{r}}^{(k_1+\ldots+k_a)} \prod_{j=1}^{a} \mathbb{E} X_{i_j}^{k_j-l_j}$, where for $j \le a$, i_j is the value of $\mathbf{r}$ on its level set $\tilde{I}_j$. This means that $\tilde{E}^{(k_1+\ldots+k_a)} = (B_{k_1+\ldots+k_a} \circ \mathbf{1}_{L(\tilde{\mathcal{I}})}) \circ (\otimes_{s=1}^{k_1+\ldots+k_a} v_s)$, where $v_s = (\mathbb{E} X_i^{k_j-l_j})_{i \le n}$ if $s \in \{\min I_1, \ldots, \min I_a\}$ and $v_s = (1, \ldots, 1)$ otherwise. Since $\|v_s\|_\infty \le (C_D L)^{k_j-l_j}$ if $s \in \{\min I_j\}_{j \le a}$ and $\|v_s\|_\infty = 1$ otherwise, by Lemma 5.1 this implies that for any $\mathcal{K} \in P_{k_1+\ldots+k_a}$,

$$\|\tilde{E}^{(k_1+\ldots+k_a)}\|_{\mathcal{K}} \le (C_D L)^{\sum_{j \le a}(k_j-l_j)} \|B_{k_1+\ldots+k_a} \circ \mathbf{1}_{L(\tilde{\mathcal{I}})}\|_{\mathcal{K}} \le C_D L^{\sum_{j \le a}(k_j-l_j)} \|B_{k_1+\ldots+k_a}\|_{\mathcal{K}}, \tag{53}$$

where in the last inequality we used Corollary 5.3.

We will now use the above inequality to prove (51). Consider the unique partition $\mathcal{K} = \{K_1, \ldots, K_b\}$ satisfying the following two conditions:

- for each $j \le b$, $J_j \subseteq K_j$,

- for each $s \in \{d+1, \ldots, k_1 + \ldots + k_a\}$ if $s \in \tilde{I}_j$ and $\pi(s) := \min \tilde{I}_j \in J_k$, then $s \in K_k$. In other words all indices s, which in the construction of $\tilde{\mathcal{I}}$ were added to I_j (i.e., elements of $\tilde{I}_j \setminus I_j$) are now added to the unique element of $\mathcal{J}$ containing $\pi(s) = \min \tilde{I}_j = \min I_j$.

Now, it is easy to see that $\|E^{(d)}\|_{\mathcal{J}} \le \|\tilde{E}^{(k_1+\ldots+k_a)}\|_{\mathcal{K}}$. Indeed, consider an arbitrary $x^{(j)} = (x_{\mathbf{r}_{J_j}}^{(j)})_{|\mathbf{r}_{J_j}| \le n}$, $j = 1, \ldots, b$, satisfying $\|x^{(j)}\|_2 \le 1$. Define $y^{(j)} = (y_{\mathbf{r}_{K_j}}^{(j)})_{|\mathbf{r}_{K_j}| \le n}$, $j = 1, \ldots, b$ with the formula

$$y_{\mathbf{r}_{K_j}}^{(j)} = x_{\mathbf{r}_{K_j \cap [d]}}^{(j)} \prod_{s \in K_j \setminus [d]} \mathbf{1}_{\{r_s = r_{\pi(s)}\}}.$$

We have $\|y^{(j)}\|_2 = \|x^{(j)}\|_2 \le 1$. Moreover, by the construction of the matrix $\tilde{E}^{(k_1+\ldots+k_a)}$ (recall (52)), we have

$$\sum_{|\mathbf{r}_{[d]}| \le n} e_{\mathbf{r}_{[d]}}^{(d)} \prod_{j=1}^{b} x_{\mathbf{r}_{J_j}}^{(j)} = \sum_{|\mathbf{r}_{[k_1+\ldots+k_a]}| \le n} \tilde{e}_{\mathbf{r}_{[k_1+\ldots+k_a]}}^{(k_1+\ldots+k_a)} \prod_{j=1}^{b} x_{\mathbf{r}_{J_j}}^{(j)} = \sum_{|\mathbf{r}_{[k_1+\ldots+k_a]}| \le n} \tilde{e}_{\mathbf{r}_{[k_1+\ldots+k_a]}}^{(k_1+\ldots+k_a)} \prod_{j=1}^{b} y_{\mathbf{r}_{K_j}}^{(j)}$$

(in the last equality we used the fact that if $\mathbf{r} \in L(\tilde{\mathcal{I}})$, then for $s > d$, $r_{\pi(s)} = r_s$ and so $y_{\mathbf{r}_{K_j}}^{(j)} = x_{\mathbf{r}_{K_j \cap [d]}}^{(j)} = x_{\mathbf{r}_{J_j}}^{(j)}$). By taking the supremum over $x^{(j)}$ one thus obtains $\|E^{(d)}\|_{\mathcal{J}} \le \|\tilde{E}^{(k_1+\ldots+k_a)}\|_{\mathcal{K}}$. Combining this inequality with (53) proves (51) and thus (50). This ends the proof of Theorem 1.4.

5.3 Application: Subgraph counting in random graphs

We will now apply results from Section 5 to some special cases of the problem of subgraph counting in Erdős-Rényi random graphs $G(n, p)$, which is often used as a test model for deviation inequalities for polynomials in independent random variables. More specifically we will investigate the problem of counting cycles of fixed length.

It turns out that Theorem 1.4 may give in some ranges of parameters optimal inequalities (leading to improvements of known results), whereas in some other regimes the estimates it gives are suboptimal.

Let us first describe the setting (we will do it in a slightly more general form that needed for our example). We will consider undirected graphs $G = (V, E)$, where V is a finite set of vertices and E is the set of edges (i.e. two-element subsets of V). By $V_G = V(G)$ and $E_G = E(G)$ we mean the set of vertices and edges (respectively) of a graph G. Also, $v_G = v(G)$ and $e_G = e(G)$ denote the number of vertices and edges in G. We say that a graph H is a subgraph of a graph G if $V_H \subseteq V_G$ and $E_H \subseteq E_G$ (thus a subgraph is non-necessarily induced). Graphs H and G are isomorphic if there is a bijection $\pi \colon V_H \to V_G$ such that for all distinct $v, w \in V_H$, $\{\pi(v), \pi(w)\} \in E_G$ iff $\{v, w\} \in E_H$.

For $p \in [0, 1]$ consider now the Erdős-Rényi random graph $G = G(n, p)$, i.e., a graph with n vertices (we will assume that $V_G = [n]$) whose edges are selected independently at random with probability p. In what follows we will be concerned with a number of copies of a given graph $H = ([k], E_H)$ in a graph G, i.e., the number of subgraphs of G which are isomorphic to H. We will denote this random variable by $Y_H(n, p)$. To relate $Y_H(n, p)$ to polynomials, let us consider the family $C(n, 2)$ of two-element subsets of $[n]$ and the family of independent random variables $X = (X_e)_{e \in C(n,2)}$, such that $\mathbb{P}(X_e = 1) = 1 - \mathbb{P}(X_e = 0) = p$ (i.e., X_e indicates whether the edge e has been selected or not). Denote moreover by $\mathrm{Aut}(H)$ the group of isomorphisms of H into itself and note that

$$Y_H(n, p) = \frac{1}{\#\mathrm{Aut}(H)} \sum_{\mathbf{i} \in [n]^{\underline{k}}} \prod_{\substack{v, w \in [k] \\ v < w, \{v, w\} \in E(H)}} X_{\{i_v, i_w\}}.$$

The right-hand side above is a homogeneous tetrahedral polynomial of degree e_H. Moreover the variables $X_{\{v, w\}}$ satisfy

$$\mathbb{E} \exp\left(X_{\{v, w\}}^2 \log(1/p) \right) = 1 - p + p \cdot \frac{1}{p} \leq 2$$

and

$$\mathbb{E} \exp\left(X_{\{v, w\}}^2 \log 2 \right) \leq 2,$$

which implies that $\|X_{\{v, w\}}\|_{\psi_2} \leq (\log(1/p))^{-1/2} \wedge (\log(2))^{-1/2} \leq \sqrt{2}(\log(2/p))^{-1/2}$.

We can thus apply Theorem 1.4 to $Y_H(n, p)$ and obtain

$$\mathbb{P}\left(|Y_H(n, p) - \mathbb{E} Y_H(n, p)| \geq t \right) \leq 2 \exp\left(-\frac{1}{C_k} \min_{1 \leq d \leq k} \min_{\mathcal{J} \in P_d} \left(\frac{t}{L_p^d \|\mathbb{E} \mathbf{D}^d f(X)\|_{\mathcal{J}}} \right)^{2/\#\mathcal{J}} \right), \quad (54)$$

where $L_p = \sqrt{2}\left(\log(2/p) \right)^{-1/2}$ and $f \colon \mathbb{R}^{C(n,2)} \to \mathbb{R}$ is given by

$$f((x_e)_{e \in C(n,2)}) = \frac{1}{\#\mathrm{Aut}(H)} \sum_{\mathbf{i} \in [n]^{\underline{k}}} \prod_{\substack{v, w \in [k] \\ v < w, \{v, w\} \in E}} x_{\{i_v, i_w\}}.$$

Deviation inequalities for subgraph counts have been studied by many authors, to mention [36, 35, 63, 34, 37, 22, 26, 25]. As it turns out the lower tail $\mathbb{P}(Y_H(n, p) \leq \mathbb{E} Y_H(n, p) - t)$ is easier than the upper tail $\mathbb{P}(Y_H(n, p) \geq \mathbb{E} Y_H(n, p) + t)$. The lower tail turns out to be also lighter than

the upper one. Since our inequalities concern $|Y_H(n,p) - \mathbb{E}Y_H(n,p)|$, we cannot hope to recover optimal lower tail estimates, however we can still hope to get bounds which in some range of parameters n, p will agree with optimal upper tail estimates.

Of particular importance in literature is the law of large numbers regime, i.e., the case when $t = \varepsilon\mathbb{E}Y_H(n,p)$. In [34] the Authors prove that for every $\varepsilon > 0$ such that $\mathbb{P}\big(Y_H(n,p) \geq (1+\varepsilon)\mathbb{E}Y_H(n,p)\big) > 0$,

$$\exp\left(-C(H,\varepsilon)M_H^*(n,p)\log\frac{1}{p}\right) \leq \mathbb{P}\big(Y_H(n,p) \geq (1+\varepsilon)\mathbb{E}Y_H(n,p)\big) \leq \exp\big(-c(H,\varepsilon)M_H^*(n,p)\big) \tag{55}$$

for certain constants $c(H,\varepsilon), C(H,\varepsilon)$ and a certain function $M_H^*(n,p)$. Since the general definition of M_H^* is rather involved we will skip the details (in the examples considered in the sequel we will provide specific formulas). Note that if one disregards the constants depending only on H and ε, the lower and upper estimate above differ by the factor $\log(1/p)$ in the exponent. To our best knowledge providing a lower and upper bound for general H, which would agree up to multiplicative constants in the exponent (depending only on H and ε, but not on n or p) is an open problem.

We will now specialize to the case when H is a cycle. For simplicity we will first present the case of the triangle K_3 (the clique with three vertices). For this graph the upper bound from [34] has been recently strengthened to match the lower one (up to a constant depending only on ε) by Chatterjee [22] and DeMarco and Kahn [26] (who also obtained a similar result for general cliques [25]). In the next section we show that if p is not too small, the inequality (54) also allows to recover the optimal upper bound. In Section 5.3.2 we provide an upper bound for cycles of arbitrary (fixed) length k, which is optimal for $p \geq n^{-\frac{k-2}{2(k-1)}}\log^{-\frac{1}{2}}n$.

5.3.1 Counting triangles

Assume that $H = K_3$ and let us analyse the behaviour of $\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}$ for $d = 1, 2, 3$. Of course in this case $\#\mathrm{Aut}(H) = 6$.

We have for any $e = \{v, w\}$, $v, w \in [n]$,

$$\frac{\partial}{\partial x_e}f(x) = \sum_{i \in [n]\setminus\{v,w\}} x_{\{i,v\}}x_{\{i,w\}}$$

and so $\|\mathbb{E}\mathbf{D}f(X)\|_{\{1\}} = (n-2)p^2\sqrt{n(n-1)/2} \leq n^2p^2$.

For $e_1 = e_2$ or when e_1 and e_2 do not have a common vertex, we have $\frac{\partial^2}{\partial x_{e_1}\partial x_{e_2}}f = 0$, whereas for e_1, e_2 sharing exactly one vertex, we have

$$\frac{\partial^2}{\partial x_{e_1}\partial x_{e_2}}f(x) = x_{\{v,w\}},$$

where v, w are the vertices of e_1, e_2 distinct from the common one. Therefore

$$\mathbb{E}\mathbf{D}^2 f(X) = p(\mathbf{1}_{\{e_1, e_2 \text{ have exactly one common vertex}\}})_{e_1,e_2 \in C(n,2)}.$$

Using the fact that $\mathbb{E}\mathbf{D}^2 f(X)$ is symmetric and for each e_1 the sum of entries of $\mathbb{E}\mathbf{D}^2 f(X)$ in the row corresponding to e_1 equals $2p(n-2)$, we obtain $\|\mathbb{E}\mathbf{D}^2 f(X)\|_{\{1\}\{2\}} = 2p(n-2) \leq 2pn$. One can also easily see that $\|\mathbb{E}\mathbf{D}^2 f(X)\|_{\{1,2\}} = p\sqrt{n(n-1)(n-2)} \leq pn^{3/2}$.

Finally

$$\frac{\partial^3}{\partial x_{e_1}\partial x_{e_2}\partial x_{e_3}}f = \mathbf{1}_{\{e_1, e_2, e_3 \text{ form a triangle}\}}$$

and thus $\|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1,2,3\}} = \sqrt{n(n-1)(n-2)} \leq n^{3/2}$. Moreover, due to symmetry we have

$$\|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1,2\}\{3\}} = \|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1,3\}\{2\}} = \|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{2,3\}\{1\}}.$$

Consider arbitrary $(x_{e_1})_{e_1\in C(n,2)}$ and $(y_{e_2,e_3})_{e_2,e_3\in C(n,2)}$ of norm one. We have

$$\sum_{e_1,e_2,e_3} \mathbf{1}_{\{e_1,\,e_2,\,e_3\text{ form a triangle}\}} x_{e_1} y_{e_2,e_3} \le \sqrt{\sum_{e_1}\Big(\sum_{e_2,e_3}\mathbf{1}_{\{e_1,\,e_2,\,e_3\text{ form a triangle}\}} y_{e_2,e_3}\Big)^2}$$

$$\le \sqrt{\sum_{e_1}\Big(\sum_{e_2,e_3}\mathbf{1}_{\{e_1,\,e_2,\,e_3\text{ form a triangle}\}}\Big)\Big(\sum_{e_2,e_3}\mathbf{1}_{\{e_1,\,e_2,\,e_3\text{ form a triangle}\}} y_{e_2,e_3}^2\Big)}$$

$$= \sqrt{2(n-2)}\sqrt{\sum_{e_2,e_3} y_{e_2,e_3}^2 \sum_{e_1}\mathbf{1}_{\{e_1,\,e_2,\,e_3\text{ form a triangle}\}}} \le \sqrt{2(n-2)},$$

where the first two inequalities follow by the Cauchy-Schwarz inequality and the last one from the fact that for each e_2, e_3 there is at most one e_1 such that e_1, e_2, e_3 form a triangle. We have thus obtained $\|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1,2\}\{3\}} = \|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1,3\}\{2\}} = \|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{2,3\}\{1\}} \le \sqrt{2n}$.

It remains to estimate $\|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1\}\{2\}\{3\}}$. For all $(x_e)_{e\in C(n,2)}$, $(y_e)_{e\in C(n,2)}$, $(z_e)_{e\in C(n,2)}$ of norm one we have by the Cauchy-Schwarz inequality

$$\sum_{e_1,e_2,e_3} \mathbf{1}_{\{e_1,\,e_2,\,e_3\text{ form a triangle}\}} x_{e_1} y_{e_2} z_{e_3} = \sum_{(i_1,i_2,i_3)\in[n]^{\underline{3}}} x_{\{i_1,i_2\}} y_{\{i_2,i_3\}} z_{\{i_1,i_3\}}$$

$$\le \sum_{i_1\in[n]}\Big(\sum_{(i_2,i_3)\in([n]\setminus\{i_1\})^{\underline{2}}} x_{\{i_1,i_2\}}^2 z_{\{i_1,i_3\}}^2\Big)^{1/2}\Big(\sum_{(i_2,i_3)\in([n]\setminus\{i_1\})^{\underline{2}}} y_{\{i_2,i_3\}}^2\Big)^{1/2}$$

$$\le \sqrt{2}\sum_{i_1\in[n]}\Big(\sum_{i_2\in[n]\setminus\{i_1\}} x_{\{i_1,i_2\}}^2\Big)^{1/2}\Big(\sum_{i_3\in[n]\setminus\{i_1\}} z_{\{i_1,i_3\}}^2\Big)^{1/2}$$

$$\le \sqrt{2}\Big(\sum_{(i_1,i_2)\in[n]^{\underline{2}}} x_{\{i_1,i_2\}}^2\Big)^{1/2}\Big(\sum_{(i_1,i_3)\in[n]^{\underline{2}}} z_{\{i_1,i_3\}}^2\Big)^{1/2} \le 2^{3/2},$$

which gives $\|\mathbb{E}\mathbf{D}^3 f(X)\|_{\{1\}\{2\}\{3\}} \le 2^{3/2}$.

Using (54) together with the above estimates, we obtain

Proposition 5.5. *For any $t > 0$,*

$$\mathbb{P}\big(|Y_{K_3}(n,p) - \mathbb{E}Y_{K_3}(n,p)| \ge t\big)$$

$$\le 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{L_p^6 n^3 + L_p^4 p^2 n^3 + L_p^2 p^4 n^4}, \frac{t}{L_p^3 n^{1/2} + L_p^2 pn}, \frac{t^{2/3}}{L_p^2}\Big)\Big),$$

where $L_p = \big(\log(2/p)\big)^{-1/2}$.

In particular for $t = \varepsilon\mathbb{E}Y_{K_3}(n,p) = \varepsilon\binom{n}{3}p^3$,

$$\mathbb{P}\big(|Y_{K_3}(n,p) - \mathbb{E}Y_{K_3}(n,p)| \ge \varepsilon\mathbb{E}Y_{K_3}(n,p)\big)$$

$$\le 2\exp\Big(-\frac{1}{C}\min\Big(\varepsilon^2 n^3 p^6 \log^3(2/p), (\varepsilon^2\wedge\varepsilon^{2/3})n^2 p^2\log(2/p)\Big)\Big).$$

Thus for $p \ge n^{-\frac{1}{4}}\log^{-\frac{1}{2}} n$ we obtain

$$\mathbb{P}\big(|Y_{K_3}(n,p) - \mathbb{E}Y_{K_3}(n,p)| \ge \varepsilon\mathbb{E}Y_{K_3}(n,p)\big) \le 2\exp\big(-(\varepsilon^2\wedge\varepsilon^{2/3})n^2 p^2\log(2/p)\big).$$

By Corollary 1.7 in [34], if $p \ge 1/n$, then $\frac{1}{C}n^2 p^2 \le M_{K_3}^*(n,p) \le Cn^2 p^2$ (recall (55)) and so for $p \ge n^{-1/4}\log^{-1/2} n$ the estimate obtained from the above proposition is optimal. As already mentioned the optimal estimate has been recently obtained in the full range of p by Chatterjee, DeMarco and Kahn. Unfortunately it seems that using our general approach we are not able

to recover the full strength of their result. From Proposition 5.5 one can also see that Theorem 1.4, when specialized to polynomials in 0-1 random variables is not directly comparable with the family of Kim-Vu inequalities. As shown in [35] (see table 2 therein), various inequalities by Kim and Vu give for the triangle counting problem exponents $-\min(n^{1/3}p^{1/6}, n^{1/2}p^{1/2})$, $-n^{3/2}p^{3/2}$, $-np$ (disregarding logarithmic factors). Thus for "large" p our inequality performs better than those by Kim-Vu, whereas for "small" p this is not the case (note that the Kim-Vu inequalities give meaningful bounds for $p \geq Cn^{-1}$ while ours only for $p \geq Cn^{-1/2}$). As already mentioned in the introduction the fact that our inequalities degenerate for small p is not surprising as even for sums of independent 0-1 random variables, when p becomes small, general inequalities for the sums of independent random variables with sub-Gaussian tails do not recover the correct tail behaviour (the $\|\cdot\|_{\psi_2}$ norm of the summands becomes much larger than the variance).

5.3.2 Counting cycles

We will now generalize Proposition 5.5 to cycles of arbitrary length. If H is a cycle of length k, then by Corollary 1.7 in [34], $\frac{1}{C}n^2p^2 \leq M_H^*(n,p) \leq Cn^2p^2$ for $p \geq 1/n$. Thus the bounds for the upper tail from (55) imply that for $p \geq 1/n$,

$$\exp\big(-C(k,\varepsilon)n^2p^2\log(1/p)\big) \leq \mathbb{P}\big(Y_H(n,p) \geq (1+\varepsilon)\mathbb{E}Y_H(n,p)\big) \leq \exp\big(-c(k,\varepsilon)n^2p^2\big)$$

for every $\varepsilon > 0$ for which the above probability is not zero.

We will show that similarly as for triangles, Theorem 1.4 allows to strengthen the upper bound if p is not too small with respect to n. More precisely, we have the following

Proposition 5.6. *Let H be a cycle of length k. Then for every $t > 0$,*

$$\mathbb{P}\big(|Y_H(n,p) - \mathbb{E}Y_H(n,p)| \geq t\big) \leq 2\exp\Big(-\frac{1}{C_k}\Big(\frac{t^2}{L_p^{2k}n^k} \wedge \min_{\substack{1 \leq l \leq d < k: \\ d < k \text{ or } l > 1}} \Big(\frac{t^{2/l}}{L_p^{2d/l}p^{2(k-d)/l}n^{(2k-d-l)/l}}\Big)\Big)\Big),$$

where $L_p = \big(\log(2/p)\big)^{-1/2}$. In particular for every $\varepsilon > 0$ and $p \geq n^{-\frac{k-2}{2(k-1)}}\log^{-1/2} n$,

$$\mathbb{P}\big(Y_H(n,p) \geq (1+\varepsilon)\mathbb{E}Y_H(n,p)\big) \leq 2\exp\Big(-\frac{1}{C_k}(\varepsilon^2 \wedge \varepsilon^{2/k})n^2p^2\log(2/p)\Big).$$

To prove the above proposition we need to estimate the corresponding $\|\cdot\|_{\mathcal{J}}$ norms. Since a major part of the argument does not rely on the fact that H is a cycle and bounds on $\|\cdot\|_{\mathcal{J}}$ norms may be of independent interest, we will now consider arbitrary graphs. Let thus H be a fixed graph with no isolated vertices.

Similarly to [34], it will be more convenient to count "ordered" copies of a graph H in $G(n,p)$. Namely, for $H = ([k], E_H)$, each sequence of k distinct vertices in the clique K_n, $\mathbf{i} \in [n]^{\underline{k}}$ determines an ordered copy $G_{\mathbf{i}}$ of H in K_n, where $G_{\mathbf{i}} = \mathbf{i}(H)$, i.e., $V(G_{\mathbf{i}}) = \mathbf{i}([k])$ and $E(G_{\mathbf{i}}) = \{\mathbf{i}(e): e \in E(H)\} = \{\{i_u, i_v\}: \{u, v\} \in E(H)\}$. Define

$$X_H(n,p) := \sum_{\mathbf{i} \in [n]^{\underline{k}}} \mathbf{1}_{\{G_{\mathbf{i}} \subseteq G(n,p)\}} = \sum_{\mathbf{i} \in [n]^{\underline{k}}} \prod_{\tilde{e} \in E(G_{\mathbf{i}})} X_{\tilde{e}}.$$

Clearly $X_H(n,p) = \#\mathrm{Aut}(H)Y_H(n,p)$ and $X_H(n,p) = f(X)$ where

$$f(x) := \sum_{\mathbf{i} \in [n]^{\underline{k}}} \prod_{\tilde{e} \in E(G_{\mathbf{i}})} x_{\tilde{e}} = \sum_{\mathbf{i} \in [n]^{\underline{k}}} \prod_{e \in E(H)} x_{\mathbf{i}(e)}. \tag{56}$$

A sequence of distinct edges $(\tilde{e}_1, \ldots, \tilde{e}_d) \in E(K_n)^{\underline{d}}$ determines a subgraph $G_0 \subseteq K_n$ with $V(G_0) = \bigcup_{i=1}^d \tilde{e}_i$, $E(G_0) = \{\tilde{e}_1, \ldots, \tilde{e}_d\}$. Note that

$$\partial_{G_0} f(x) := \frac{\partial^d f(x)}{\partial x_{\tilde{e}_1} \cdots \partial x_{\tilde{e}_d}} = \sum_{\mathbf{i} \in [n]^{\underline{k}}: G_{\mathbf{i}} \supseteq G_0} \prod_{\tilde{e} \in E(G_{\mathbf{i}}) \setminus E(G_0)} x_{\tilde{e}}$$

31

and thus

$$\mathbb{E}\partial_{G_0}f(X) = p^{e(H)-d}\#\{\mathbf{i} \in [n]^{\underline{k}}\colon G_0 \subseteq G_{\mathbf{i}}\}$$

Consider $\mathbf{e} = (e_1, \dots, e_d) \in E(H)^{\underline{d}}$ and let $H_0(\mathbf{e})$ be the subgraph of H with $V(H_0(\mathbf{e})) = \bigcup_{i=1}^d e_i$, $E(H_0(\mathbf{e})) = \{e_1, \dots, e_d\}$. Clearly, for any $\mathbf{i} \in [n]^{\underline{k}}$, $\mathbf{i}(H_0(\mathbf{e})) \subseteq G_{\mathbf{i}}$. We write $(e_1, \dots e_d) \simeq (\tilde{e}_1, \dots, \tilde{e}_d)$ if there exists $\mathbf{i} \in [n]^{\underline{k}}$ such that $\mathbf{i}(e_j) = \tilde{e}_j$ for $j = 1, \dots, d$. Note that given $(\tilde{e}_1, \dots, \tilde{e}_d) \in E(K_n)^{\underline{d}}$ and the corresponding graph G_0,

$$\#\{\mathbf{i} \in [n]^{\underline{k}}\colon G_0 \subseteq G_{\mathbf{i}}\} = \sum_{\mathbf{e} \in E(H)^{\underline{d}}} \#\{\mathbf{i} \in [n]^{\underline{k}}\colon \mathbf{i}(e_j) = \tilde{e}_j \text{ for } j = 1, \dots, d\}$$

$$= \sum_{\mathbf{e} \in E(H)^{\underline{d}}} 2^{s(H_0(\mathbf{e}))}(n - v(H_0(\mathbf{e})))^{\underline{k-v(H_0(\mathbf{e}))}}\mathbf{1}_{\{(\tilde{e}_1, \dots, \tilde{e}_d) \simeq \mathbf{e}\}},$$

where for a graph G, $v(G)$ is the number of vertices of G and $s(G)$ is the number of edges in G with no other adjacent edge. Therefore,

$$\mathbb{E}\mathbf{D}^d f(X) = p^{e(H)-d} \sum_{\mathbf{e} \in E(H)^{\underline{d}}} 2^{s(H_0(\mathbf{e}))}(n - v(H_0(\mathbf{e})))^{\underline{k-v_{H_0(\mathbf{e})}}}\left(\mathbf{1}_{\{(\tilde{e}_1, \dots, \tilde{e}_d) \simeq \mathbf{e}\}}\right)_{(\tilde{e}_1 \dots, \tilde{e}_d)}.$$

Let $\mathcal{J}$ be a partition of $[d]$. By the triangle inequality for the norms $\|\cdot\|_{\mathcal{J}}$,

$$\left\|\mathbb{E}\mathbf{D}^d f(X)\right\|_{\mathcal{J}} \le p^{e(H)-d} \sum_{\mathbf{e} \in E(H)^{\underline{d}}} 2^{s(H_0(\mathbf{e}))}n^{k-v(H_0(\mathbf{e}))} \left\|\left(\mathbf{1}_{\{(\tilde{e}_1, \dots, \tilde{e}_d) \simeq \mathbf{e}\}}\right)_{(\tilde{e}_1 \dots, \tilde{e}_d)}\right\|_{\mathcal{J}}. \quad (57)$$

The norms appearing on the right hand side of (57) are handled by the following

Lemma 5.7. *Fix* $1 \le d \le e(H)$, $\mathbf{e} = (e_1, \dots, e_d) \in E(H)^{\underline{d}}$ *and* $\mathcal{J} = \{J_1, \dots, J_l\} \in P_d$. *Let* $H_0 = H_0(\mathbf{e})$ *and for* $r = 1, \dots, l$, *let* H_r *be a subgraph of* H_0 *spanned by the set of edges* $\{e_j \colon j \in J_r\}$. *Then,*

$$\left\|\left(\mathbf{1}_{\{(\tilde{e}_1, \dots, \tilde{e}_d) \simeq (e_1, \dots, e_d)\}}\right)_{(\tilde{e}_1 \dots, \tilde{e}_d)}\right\|_{\mathcal{J}} \le 2^{-s(H_0)+\frac{1}{2}\sum_{r=1}^l s(H_r)}$$

$$\times n^{\frac{1}{2}\#\{v \in V(H_0)\colon v \in V(H_r) \text{ for exactly one } r \in [l]\}}.$$

Proof. We shall bound the sum

$$\sum_{\tilde{e}_1, \dots, \tilde{e}_d \in E(K_n)} \mathbf{1}_{\{(\tilde{e}_1, \dots, \tilde{e}_d) \simeq \mathbf{e}\}} \prod_{r=1}^l x^{(r)}_{(\tilde{e}_j)_{j \in J_r}} \quad (58)$$

under the constraints $\sum_{(\tilde{e}_j)_{j \in J_r} \in E(K_n)^{J_r}} \left(x^{(r)}_{(\tilde{e}_j)_{j \in J_r}}\right)^2 \le 1$ for $r = 1, \dots, l$. Note that we can assume $x^{(r)} \ge 0$ for all $r \in [l]$. Rewrite the sum (58) as the sum over a sequence of vertices instead of edges:

$$2^{-s(H_0)} \sum_{\mathbf{i} \in [n]^{\underline{V(H_0)}}} \prod_{r=1}^l x^{(r)}_{(\mathbf{i}(e_j))_{j \in J_r}},$$

where for two sets A, B, $A^{\underline{B}}$ is the set of 1-1 functions from B to A. Further note that it is enough to prove the desired bound for the sum

$$2^{-s(H_0)} \sum_{\mathbf{i} \in [n]^{\underline{V(H_0)}}} \prod_{r=1}^l y^{(r)}_{\mathbf{i}_{V(H_r)}} \quad (59)$$

32

under the constraints $2^{-s(H_r)} \sum_{\mathbf{i} \in [n]^{\underline{V(H_r)}}} \left(y^{(r)}_{\mathbf{i}_{V(H_r)}} \right)^2 \leq 1$ for each $r = 1, \ldots, l$. Indeed, given x's, for each $r = 1, \ldots, l$ and all $\mathbf{i} \in [n]^{\underline{V(H_r)}}$ take $y^{(r)}_{\mathbf{i}_{V(H_r)}} = x^{(r)}_{(\mathbf{i}(e_j)_{j \in J_r})}$ and notice that the sum (59) equals the sum (58) while the constraints for x's imply the constraints for y's. Finally, by homogeneity and the fact that the sum (59) does not depend on the full graph structure but only on the sets of vertices of the graphs H_r, the lemma will follow from the statement: For a sequence of finite, non-empty sets $V_1, \ldots, V_l$, let $V = V_1 \cup \ldots \cup V_l$. Then

$$\sum_{\mathbf{i} \in [n]^{\underline{V}}} \prod_{r=1}^{l} y^{(r)}_{\mathbf{i}_{V_r}} \leq n^{\frac{1}{2} \#\{v \in V \,:\, v \in V_r \text{ for exactly one } r \in [l]\}} \tag{60}$$

for $y^{(1)}, \ldots, y^{(l)} \geq 0$ satisfying

$$\sum_{\mathbf{i} \in [n]^{\underline{V_r}}} \left(y^{(r)}_{\mathbf{i}_{V_r}} \right)^2 \leq 1. \tag{61}$$

We prove (60) by induction on $\#V$. For $V = \emptyset$ (and $l = 0$), (60) holds trivially. For the induction step fix any $v_0 \in V$ and put $R = \{r \in [l] \colon v_0 \in V_r\}$. We write

$$\sum_{\mathbf{i} \in [n]^{\underline{V}}} \prod_{r=1}^{l} y^{(r)}_{\mathbf{i}_{V_r}} = \sum_{\mathbf{i} \in [n]^{\underline{V \setminus \{v_0\}}}} \left(\left(\prod_{r \in [l] \setminus R} y^{(r)}_{\mathbf{i}_{V_r}} \right) \sum_{i_{v_0} \in [n] \setminus \mathbf{i}(V \setminus \{v_0\})} \prod_{r \in R} y^{(r)}_{\mathbf{i}_{V_r}} \right).$$

We bound the inner sum using the Cauchy-Schwarz inequality. If $\#R \geq 2$, we get

$$\sum_{i_{v_0} \in [n] \setminus \mathbf{i}(V \setminus \{v_0\})} \prod_{r \in R} y^{(r)}_{\mathbf{i}_{V_r}} \leq \prod_{r \in R} \left(\sum_{i_{v_0} \in [n] \setminus \mathbf{i}(V \setminus \{v_0\})} \left(y^{(r)}_{\mathbf{i}_{V_r}} \right)^2 \right)^{1/2},$$

and if $R = \{r_0\}$ then

$$\sum_{i_{v_0} \in [n] \setminus \mathbf{i}(V \setminus \{v_0\})} y^{(r_0)}_{\mathbf{i}_{V_{r_0}}} \leq \sqrt{n} \left(\sum_{i_{v_0} \in [n] \setminus \mathbf{i}(V \setminus \{v_0\})} \left(y^{(r_0)}_{\mathbf{i}_{V_{r_0}}} \right)^2 \right)^{1/2}.$$

Now, for each $r \in R$ put $W_r = V_r \setminus \{v_0\}$ and define

$$z^{(r)}_{\mathbf{i}_{W_r}} = \left(\sum_{i_{v_0} \in [n] \setminus \mathbf{i}(W_r)} \left(y^{(r)}_{\mathbf{i}_{V_r}} \right)^2 \right)^{1/2} \quad \text{for all } \mathbf{i}_{W_r} \in [n]^{\underline{W_r}}.$$

Note that if $W_r = \emptyset$ then $z^{(r)}$ is a scalar and by (61), $0 \leq z^{(r)} \leq 1$. For $r \in [l] \setminus R$, just put $W_r = V_r$ and $z^{(r)} \equiv y^{(r)}$. Let $L = \{r \in [l] \colon W_r \neq \emptyset\}$. Combining the estimates obtained above, we arrive at

$$\sum_{\mathbf{i} \in [n]^{\underline{V}}} \prod_{r=1}^{l} y^{(r)}_{\mathbf{i}_{V_r}} \leq (\sqrt{n})^{\mathbf{1}_{\{v_0 \in V_r \text{ for exactly one } r \in [l]\}}} \sum_{\mathbf{i} \in [n]^{\underline{V \setminus \{v_0\}}}} \prod_{r \in L} z^{(r)}_{\mathbf{i}_{W_r}}.$$

Now we use the induction hypothesis for the sequence of sets $(W_r)_{r \in L}$ and the vectors $z^{(r)}$, $r \in L$ (note that $\sum_{\mathbf{i} \in [n]^{\underline{W_r}}} (z^{(r)}_{\mathbf{i}_{W_r}})^2 \leq 1$). $\qquad \square$

Remark The bound in Lemma 5.7 is essentially optimal, at least for large n, say $n \geq 2k$. To see this let us analyse optimality of (60) under the constraints (61) (it is easy to see that this is equivalent to the optimality in the original problem). Denote $V_0 = \{v \in V \colon v \in V_r$ for exactly one $r \in [l]\}$. Fix any $\mathbf{i}^{(0)} \in [n]^{\underline{k}}$. Then for $r = 1, \ldots, l$ take

$$
y_{\mathbf{i}_{V_r}}^{(r)} = \begin{cases} n^{-\frac{1}{2}\#(V_r \cap V_0)} & \text{if } \mathbf{i}_{V_r \setminus V_0} \equiv \mathbf{i}_{V_r \setminus V_0}^{(0)} \\ 0 & \text{otherwise.} \end{cases}
$$

The vectors $y^{(r)}$ satisfy the constraints (61) and

$$
\sum_{\mathbf{i} \in [n]^{\underline{V}}} \prod_{r=1}^{l} y_{\mathbf{i}_{V_r}}^{(r)} = \sum_{\mathbf{i} \in [n]^{\underline{V}} \colon \mathbf{i}_{V \setminus V_0} \equiv \mathbf{i}_{V \setminus V_0}^{(0)}} \prod_{r=1}^{l} n^{-\frac{1}{2}\#(V_r \cap V_0)}
$$

$$
= (n - \#(V \setminus V_0))^{\underline{\#V_0}} \, n^{-\frac{1}{2}\#V_0} \geq (n/2)^{\#V_0} n^{-\frac{1}{2}\#V_0} = 2^{-\#V_0} n^{\frac{1}{2}\#V_0}.
$$

Combining Lemma 5.7 with (57) we obtain

Lemma 5.8. *Let H be any graph with k vertices, which are not isolated, and let f be defined by (56). Then for any $1 \leq d \leq e(H)$ and any $\mathcal{J} = \{J_1, \ldots, J_l\} \in P_d$,*

$$
\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}
$$
$$
\leq p^{e(H)-d} \sum_{\mathbf{e} \in E(H)^{\underline{d}}} 2^{\frac{1}{2}\sum_{r=1}^{l} s(H_r(\mathbf{e}))} n^{k - v(H_0(\mathbf{e})) + \frac{1}{2}\#\{v \in V(H_0(\mathbf{e})) \colon v \in V(H_r(\mathbf{e})) \text{ for exactly one } r \in [l]\}},
$$

where for $\mathbf{e} \in E(H)^{\underline{d}}$ and $r \in [l]$, $H_r(\mathbf{e})$ is the subgraph of $H_0(\mathbf{e})$ spanned by $\{e_j \colon j \in J_r\}$.

We are now ready for

Proof of Proposition 5.6. We will use Lemma 5.8 to estimate $\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}}$ for any $d \leq k$ and $\mathcal{J} \in P_d$ with $\#\mathcal{J} = l$. Note that for any $\mathbf{e} \in \mathbb{E}(H)^{\underline{d}}$,

$$
v(H_0(\mathbf{e})) - \frac{1}{2}\#\{v \in V(H_0(\mathbf{e})) \colon v \in V(H_r(\mathbf{e})) \text{ for exactly one } r \in [l]\}
$$

$$
= \frac{1}{2}\big(v(H_0(\mathbf{e})) + \#\{v \in V(H_0(\mathbf{e})) \colon v \text{ belongs to more than one } V(H_r(\mathbf{e}))\}\big)
$$

$$
\begin{cases} = k/2 & \text{if } d = k \text{ and } l = 1, \\ \geq \frac{1}{2}(d+l) & \text{otherwise,} \end{cases}
$$

where to get the second inequality we used the fact that each vertex of H has degree two and the inclusion-exclusion formula. Thus we obtain

$$
\|\mathbb{E}\mathbf{D}^k f(X)\|_{\{[k]\}} \leq n^{k/2},
$$
$$
\|\mathbb{E}\mathbf{D}^d f(X)\|_{\mathcal{J}} \leq C_k p^{k-d} n^{k - \frac{1}{2}d - \frac{1}{2}l} \quad \text{if } d < k \text{ or } l > 1.
$$

Together with (54) this yields the first inequality of the proposition. Using the fact that $\mathbb{E}Y_H(n,p) \geq \frac{1}{C_k} n^k p^k$, the second inequality follows by simple calculations. $\square$

6 Refined inequalities for polynomials in independent random variables satisfying the modified log-Sobolev inequality

In this section we refine the inequalities which can be obtained from Theorem 3.3 for polynomials in independent random variables satisfying the β-modified log-Sobolev inequality (15) with $\beta > 2$. To this end we will use Theorem 3.4 together with a result from [2], which is a counterpart of Theorem 3.1 for homogeneous tetrahedral polynomials in general independent symmetric random variables with log-concave tails, however only of degree at most 3. We recall that for a set I, by P_I we denote the family of partitions of I into pairwise disjoint, nonempty sets.

Theorems 3.1 and 3.2 and 3.4 from [2] specialized to Weibull variables can be translated into

Theorem 6.1. *Let $\alpha \in [1, 2]$ and let $Y_1, \ldots, Y_n$ be a sequence of i.i.d. symmetric random variables satisfying $\mathbb{P}(|Y_i| \geq t) = \exp(-t^\alpha)$. Define $Y = (Y_1, \ldots, Y_n)$ and let $Z_1, \ldots, Z_d$ be independent copies of Y. Consider a d-indexed matrix A. Define also*

$$m_d(p, A) = \sum_{I \subseteq [d]} \sum_{\mathcal{J} \in P_I} \sum_{\mathcal{K} \in P_{[d] \setminus I}} p^{\#\mathcal{J}/2 + \#\mathcal{K}/\alpha} \|A\|_{\mathcal{J}|\mathcal{K}}, \tag{62}$$

where for $\mathcal{J} = \{J_1, \ldots, J_r\} \in P_I$ and $\mathcal{K} = \{K_1, \ldots, K_k\} \in P_{[d] \setminus I}$,

$$\|A\|_{\mathcal{J}|\mathcal{K}} = \sum_{s_1 \in K_1, \ldots, s_k \in K_k} \sup \Big\{ \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} \prod_{l=1}^r x_{\mathbf{i}_{J_l}}^{(l)} \prod_{l=1}^k y_{\mathbf{i}_{J_l}}^{(l)} \colon \|(x_{\mathbf{i}_{J_l}}^{(l)})\|_2 \leq 1, \text{ for } 1 \leq l \leq r,$$
$$\sum_{i_{s_l} \leq n} \|(y_{\mathbf{i}_{K_l}}^{(l)})_{\mathbf{i}_{K_l} \setminus \{s_l\}}\|_2^\alpha \leq 1, \text{ for } 1 \leq l \leq k \Big\}.$$

If $d \leq 3$, then for any $p \geq 2$,

$$C_d^{-1} m_d(p, A) \leq \|\langle A, Z_1 \otimes \cdots \otimes Z_d \rangle\|_p \leq C_d m_d(p, A).$$

Moreover, if $\alpha = 1$, then the above inequality holds for all $d \geq 1$.

Before we proceed, let us provide a few specific examples of the norms $\|A\|_{\mathcal{J}|\mathcal{K}}$, which for $\alpha < 2$ are more complicated than in the Gaussian case. In what follows, $\beta = \frac{\alpha}{\alpha - 1}$ (with $\beta = \infty$ for $\alpha = 1$). For $d = 1$,

$$\|(a_i)\|_{\{1\}|\emptyset} = \sup \Big\{ \sum a_i x_i \colon \sum x_i^2 \leq 1 \Big\} = |(a_i)|_2,$$
$$\|(a_i)\|_{\emptyset|\{1\}} = \sup \Big\{ \sum a_i y_i \colon \sum |y_i|^\alpha \leq 1 \Big\} = |(a_i)|_\beta.$$

For $d = 2$, $\|(a_{ij})\|_{\{1,2\}|\emptyset} = \|(a_{ij})\|_{\mathrm{HS}}$, $\|(a_{ij})\|_{\{1\}\{2\}|\emptyset} = \|(a_{ij})\|_{\ell_2 \to \ell_2}$,

$$\|(a_{ij})\|_{\{1\}|\{2\}} = \sup \Big\{ \sum a_{ij} x_i y_j \colon \sum x_i^2 \leq 1, \sum |y_j|^\alpha \leq 1 \Big\} = \|(a_{ij})\|_{\ell_\alpha \to \ell_2},$$
$$\|(a_{ij})\|_{\{2\}|\{1\}} = \sup \Big\{ \sum a_{ij} y_i x_j \colon \sum x_j^2 \leq 1, \sum |y_i|^\alpha \leq 1 \Big\} = \|(a_{ij})\|_{\ell_2 \to \ell_\beta},$$
$$\|(a_{ij})\|_{\emptyset|\{1\}\{2\}} = \sup \Big\{ \sum a_{ij} y_i z_j \colon \sum |y_i|^\alpha \leq 1, \sum |z_j|^\alpha \leq 1 \Big\} = \|(a_{ij})\|_{\ell_\alpha \to \ell_\beta},$$

and

$$\|(a_{ij})\|_{\emptyset|\{1,2\}} = \sup \Big\{ \sum a_{ij} y_{ij} \colon \sum_i \big(\sum_j y_{ij}^2 \big)^{\frac{\alpha}{2}} \leq 1 \Big\} + \sup \Big\{ \sum a_{ij} y_{ij} \colon \sum_j \big(\sum_i y_{ij}^2 \big)^{\frac{\alpha}{2}} \leq 1 \Big\}$$
$$= \Big(\sum_i \big(\sum_j a_{ij}^2 \big)^{\beta/2} \Big)^{1/\beta} + \Big(\sum_j \big(\sum_i a_{ij}^2 \big)^{\beta/2} \Big)^{1/\beta}.$$

For $d = 3$, we have, for example,

$$\|(a_{ijk})\|_{\{2\}|\{1\}\{3\}} = \sup\Big\{\sum a_{ijk}y_i x_j z_k \colon \sum |x_j|^2 \le 1, \sum |y_i|^\alpha \le 1, \sum |z_k|^\alpha \le 1\Big\},$$

$$\|(a_{ijk})\|_{\{2\}|\{1,3\}} = \sup\Big\{\sum a_{ijk}x_j y_{ik} \colon \sum x_j^2 \le 1, \sum_i \Big(\sum_k y_{ik}^2\Big)^{\frac{\alpha}{2}} \le 1\Big\}$$

$$+ \sup\Big\{\sum a_{ijk}x_j y_{ik} \colon \sum x_j^2 \le 1, \sum_k \Big(\sum_i y_{ik}^2\Big)^{\frac{\alpha}{2}} \le 1\Big\},$$

$$\|(a_{ijk})\|_{\emptyset|\{1\}\{2,3\}} = \sup\Big\{\sum a_{ijk}y_i z_{jk} \colon \sum |y_i|^\alpha \le 1, \sum_j \Big(\sum_k z_{jk}^2\Big)^{\frac{\alpha}{2}} \le 1\Big\}$$

$$+ \sup\Big\{\sum a_{ijk}y_i z_{jk} \colon \sum |y_i|^\alpha \le 1, \sum_k \Big(\sum_j z_{jk}^2\Big)^{\frac{\alpha}{2}} \le 1\Big\}.$$

In particular, from Theorem 6.1 it follows that for $\alpha \in [1,2]$, if $Y = (Y_1, \ldots, Y_n)$ is as in Theorem 6.1 then for every $x \in \mathbb{R}^n$,

$$\frac{1}{C}(\sqrt{p}|x|_2 + p^{1/\alpha}|x|_\beta) \le \|\langle x, Y \rangle\|_p \le C(\sqrt{p}|x|_2 + p^{1/\alpha}|x|_\beta),$$

where $|\cdot|_r$ stands for ℓ_r^n norm (see also [28]). Thus, for $\beta \in (2, \infty)$, the inequality of Theorem 3.4, for $m = n$, $k = 1$ and a $\mathcal{C}^1$ function $f \colon \mathbb{R}^n \to \mathbb{R}$, can be written in the form

$$\|f(X) - \mathbb{E}f(X)\|_p \le C_\beta \|\langle \nabla f(X), Y \rangle\|_p. \tag{63}$$

This allows for induction, just as in the proof of Proposition 3.2, except that instead of Gaussian vectors we will have independent copies of Y. We can thus repeat the proof of Theorem 3.3, using the above observation and Theorem 6.1 instead of Theorem 3.1. This argument will then yield the following proposition, which is a counterpart of Theorem 3.3. At the moment we can prove it only for $D \le 3$, clearly generalizing Theorem 6.1 to chaoses of arbitrary degree would immediately imply it for general D.

Proposition 6.2. *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$, with independent components. Let $\beta \in (2, \infty)$ and assume that for all $i \le n$, X_i satisfies the β-modified logarithmic Sobolev inequality with constant D_{LS_β}. Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^D$ function. Define*

$$m(p, f) = \big\|m_D(p, \mathbf{D}^D f(X))\big\|_p + \sum_{1 \le d \le D-1} m_d(p, \mathbb{E}\mathbf{D}^d f(X)),$$

where $m_d(p, A)$ is defined by (62) with $\alpha = \frac{\beta}{\beta-1}$. If $D \le 3$ then for $p \ge 2$,

$$\|f(X) - \mathbb{E}f(X)\|_p \le C_{\beta, D_{LS_\beta}} m(p, f).$$

As a consequence, for all $p \ge 2$,

$$\mathbb{P}\big(|f(X) - \mathbb{E}f(X)| \ge C_{\beta, D_{LS_\beta}} m(p, f)\big) \le e^{-p}.$$

Remarks

1. For $\beta = 2$, the estimates of the above proposition agree with those of Theorem 1.2. For $\beta > 2$ it improves on what can be obtained from Theorem 3.3 in two aspects (of course just for $D \le 3$). First, the exponent of p is smaller as $(\gamma - 1/2)d + \#(\mathcal{J} \cup \mathcal{K})/2 = (1/\alpha - 1/2)d + \#\mathcal{J}/2 + \#\mathcal{K}/2 \ge \#\mathcal{J}/2 + \#\mathcal{K}/\alpha$. Second $\|A\|_{\mathcal{J} \cup \mathcal{K}} \ge \|A\|_{\mathcal{J}|\mathcal{K}}$ (since for $\alpha < 2$, $|x|_\alpha \ge |x|_2$, so the supremum on the left hand side is taken over a larger set).

2. From results in [2] it follows that if f is a tetrahedral polynomial of degree D and X_i are i.i.d. symmetric random variables satisfying $\mathbb{P}(|X_i| \geq t) = \exp(-t^\alpha)$, then the inequalities of Proposition 6.2 can be reversed (up to constants), i.e.,

$$\|f(X) - \mathbb{E}f(X)\|_p \geq \frac{1}{C_D} m_f(p).$$

This is true for any positive integer D.

3. One can also consider another functional inequality, which may be regarded a counterpart of (15) for $\beta = \infty$. We say that a random vector X in $\mathbb{R}^n$ satisfies the Bobkov-Ledoux inequality if for all locally Lipschitz positive functions such that $|\nabla f(x)|_\infty := \max_{1 \leq i \leq n} |\frac{\partial}{\partial x_i} f(x)| \leq d_{BL} f(x)$ for all x,

$$\mathrm{Ent} f^2(X) \leq D_{BL} \mathbb{E}|\nabla f(X)|^2. \tag{64}$$

This inequality has been introduced in [9] to provide a simple proof of Talagrand's two-level concentration for the symmetric exponential measure in $\mathbb{R}^n$. Here $|\frac{\partial}{\partial x_i} f(x)|$ is defined as "partial length of gradient" (see (12)). Thus in the case of differentiable functions $|\nabla f|_\infty$ coincides with the ℓ_∞^n norm of the "true" gradient.

In view of Theorem 3.4 it is natural to conjecture that the Bobkov-Ledoux inequality implies

$$\|f(X) - \mathbb{E}f(X)\|_p \leq C\left(\sqrt{p}\big\||\nabla f(X)|\big\|_p + p\big\||\nabla f(X)|_\infty\big\|_p\right), \tag{65}$$

which in turn implies (63) with $Y = (Y_1, \ldots, Y_n)$ being a vector of independent symmetric exponential variables and some $C_\infty < \infty$. This would yield an analogue of Proposition 6.2 for $\beta = \infty$, this time with no restriction on D.

Unfortunately at present we do not know whether the implication (64) $\implies$ (65) holds true or even if (65) holds for the symmetric exponential measure in $\mathbb{R}^n$. We only are able to prove the following weaker inequality, which is however not sufficient to obtain a counterpart of Proposition 6.2 for $\beta = \infty$.

Proposition 6.3. *If X is a random vector in $\mathbb{R}^n$, which satisfies (64), then for any locally Lipschitz function $f \colon \mathbb{R}^n \to \mathbb{R}$, and any $p \geq 2$,*

$$\|f(X) - \mathbb{E}f(X)\|_p \leq 3\left(D_{BL}^{1/2}\sqrt{p}\big\||\nabla f(X)|\big\|_p + d_{BL}^{-1}p\big\||\nabla f(X)|_\infty\big\|_\infty\right).$$

Proof. To simplify the notation we suppress the argument X. In what follows $\|\cdot\|_p$ denotes the L_p norm with respect to the distribution of X.

Let us fix $p \geq 2$ and consider $f_1 = \max(f, \|f\|_p/2)$. We have

$$\|f_1\|_p \geq \|f\|_p, \tag{66}$$

$$\|f_1\|_2 \leq \frac{1}{2}\|f\|_p + \|f\|_2,$$

$$\|f_1\|_p \leq \frac{3}{2}\|f\|_p \leq 3\min f_1.$$

Moreover, f_1 is locally Lipschitz and we have pointwise estimates $|\nabla f_1| \leq |\nabla f|$, $|\nabla f_1|_\infty \leq |\nabla f|_\infty$. Assume now that we have proved that

$$\|f_1\|_p \leq \|f_1\|_2 + \sqrt{\frac{D_{BL}}{2}}\sqrt{p}\big\||\nabla f_1|\big\|_p + \frac{3p}{2d_{BL}}\big\||\nabla f_1|_\infty\big\|_\infty. \tag{67}$$

Then, together with the two first inequalities of (66), it yields

$$\|f\|_p \le \|f_1\|_p \le \|f_1\|_2 + \sqrt{\frac{D_{BL}}{2}}\sqrt{p}\big\||\nabla f_1|\big\|_p + \frac{3p}{2d_{BL}}\big\||\nabla f_1|_\infty\big\|_\infty$$

$$\le \frac{1}{2}\|f\|_p + \|f\|_2 + \sqrt{\frac{D_{BL}}{2}}\sqrt{p}\big\||\nabla f|\big\|_p + \frac{3p}{2d_{BL}}\big\||\nabla f|_\infty\big\|_\infty,$$

which gives

$$\|f\|_p \le 2\Big(\|f\|_2 + \sqrt{\frac{D_{BL}}{2}}\sqrt{p}\big\||\nabla f|\big\|_p + \frac{3p}{2d_{BL}}\big\||\nabla f|_\infty\big\|_\infty\Big). \tag{68}$$

Since (64) implies the Poincaré inequality with constant $D_{BL}/2$ (see e.g. Proposition 2.3 in [27]), we can conclude the proof applying (68) to $|f - \mathbb{E}f|$ (similarly as in the proof of Theorem 3.4). Thus it is enough to prove (67).

From now on we are going to work with the function f_1 only, so for brevity we will drop the subscript and write f instead of f_1. Assume $\|f\|_p \ge \frac{3p}{2d_{BL}}\||\nabla f|_\infty\|_\infty$ (otherwise (67) is trivially satisfied). Then, using the third inequality of (66), for $2 \le t \le p$ and all $x \in \mathbb{R}^n$,

$$|\nabla f^{t/2}(x)|_\infty \le \frac{t}{2}f^{t/2-1}(x)|\nabla f(x)|_\infty \le \frac{3}{2}f^{t/2}(x)\frac{p|\nabla f(x)|_\infty}{\|f\|_p} \le d_{BL}f^{t/2}(x).$$

We can thus apply (64) with $f^{t/2}$, which together with Hölder's inequality gives

$$\mathrm{Ent}f^t \le D_{BL}\mathbb{E}|\nabla f^{t/2}|^2 \le D_{BL}\frac{t^2}{4}\mathbb{E}\big(f^{t-2}|\nabla f|^2\big) \le D_{BL}\frac{t^2}{4}\big\||\nabla f|\big\|_t^2(\mathbb{E}f^t)^{1-\frac{2}{t}}.$$

Now, as in the proof of Theorem 3.4, we have

$$\frac{d}{dt}(\mathbb{E}f^t)^{2/t} = \frac{2}{t^2}(\mathbb{E}f^t)^{\frac{2}{t}-1}\mathrm{Ent}f^t \le \frac{D_{BL}}{2}\big\||\nabla f|\big\|_p^2,$$

which upon integrating gives

$$\|f\|_p^2 \le \|f\|_2^2 + \frac{D_{BL}}{2}p\big\||\nabla f|\big\|_p^2,$$

which clearly implies (67). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

7 Appendix

7.1 Decoupling inequalities

Let us here state the main decoupling result for U-statistics (Theorem 1 in [24]).

Theorem 7.1. *For natural numbers $n \ge d$ let $(X_i)_{i=1}^n$ be a sequence of independent random variables with values in a measurable space $(S, \mathcal{S})$ and let $(X_i^{(j)})_{i=1}^n$ $j = 1, \ldots, d$ be d independent copies of this sequence. Let B be a separable Banach space and for each $\mathbf{i} \in [n]^{\underline{d}}$ let $h_\mathbf{i} \colon S^d \to B$ be a measurable function. Then for all $t > 0$,*

$$\mathbb{P}\Big(\big\|\sum_{\mathbf{i}\in[n]^{\underline{d}}} h_\mathbf{i}(X_{i_1},\ldots,X_{i_d})\big\| > t\Big) \le C_d\mathbb{P}\Big(\big\|\sum_{\mathbf{i}\in[n]^{\underline{d}}} h_\mathbf{i}(X_{i_1}^{(1)},\ldots,X_{i_d}^{(d)})\big\| > t/C_d\Big).$$

In consequence for all $p \ge 1$,

$$\Big\|\sum_{\mathbf{i}\in[n]^{\underline{d}}} h_\mathbf{i}(X_{i_1},\ldots,X_{i_d})\Big\|_p \le C_d\Big\|\sum_{\mathbf{i}\in[n]^{\underline{d}}} h_\mathbf{i}(X_{i_1}^{(1)},\ldots,X_{i_d}^{(d)})\Big\|_p.$$

If moreover the functions $h_{\mathbf{i}}$ are symmetric in the sense that, for all $x_1, \ldots, x_d \in S$ and all permutations $\pi \colon [d] \to [d]$, $h_{i_1, \ldots, i_d}(x_1, \ldots, x_d) = h_{i_{\pi_1}, \ldots, i_{\pi_d}}(x_{\pi_1}, \ldots, x_{\pi_d})$, then for all $t > 0$,

$$\mathbb{P}\Big(\Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} h_{\mathbf{i}}(X_{i_1}^{(1)}, \ldots, X_{i_d}^{(d)})\Big\| > t\Big) \le C_d \mathbb{P}\Big(\Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} h_{\mathbf{i}}(X_{i_1}, \ldots, X_{i_d})\Big\| > t/C_d\Big)$$

and in consequence for all $p \ge 1$,

$$\Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} h_{\mathbf{i}}(X_{i_1}^{(1)}, \ldots, X_{i_d}^{(d)})\Big\|_p \le C_d \Big\| \sum_{\mathbf{i} \in [n]^{\underline{d}}} h_{\mathbf{i}}(X_{i_1}, \ldots, X_{i_d})\Big\|_p.$$

7.2 Proof of Lemma 5.4

Without loss of generality we can assume that $M = 1$. It is easy to see that for some constant C_k and $t > 1$, $\mathbb{P}(C_k|g_{i1} \cdots g_{ik}| > t) \ge 2\exp(-t^{2/k})$. Since $\mathbb{P}(|Y_i| \ge t) \le 2\exp(-t^{2/k})$, we get

$$\mathbb{P}(|Y_i \mathbf{1}_{|Y_i| \ge 1}| \ge t) \le \mathbb{P}(C_k |g_{i1} \cdots g_{ik}| > t).$$

Therefore, using the inverse of the distribution function, we can define i.i.d. copies $\tilde{Y}_i$ of $|Y_i \mathbf{1}_{|Y_i| \ge 1}|$ and i.i.d copies Z_i of $|g_{i1} \cdots g_{ik}|$, such that $\tilde{Y}_i \le C_k Z_i$ pointwise. We may assume that these copies are defined on a common probability space with Y_i and g_{ij}. We can now write for a sequence ε_i of i.i.d. Rademacher variables independent of all he variables introduced so far.

$$\begin{aligned}
\|\sum_{i=1}^{n} a_i Y_i\|_p &= \|\sum_{i=1}^{n} a_i \varepsilon_i |Y_i|\|_p \\
&\le \|\sum_{i=1}^{n} a_i \varepsilon_i |Y_i \mathbf{1}_{\{|Y_i|<1\}}|\|_p + \|\sum_{i=1}^{n} a_i \varepsilon_i |Y_i \mathbf{1}_{\{|Y_i| \ge 1\}}|\|_p \\
&= \|\sum_{i=1}^{n} a_i \varepsilon_i |Y_i \mathbf{1}_{\{|Y_i|<1\}}|\|_p + \|\sum_{i=1}^{n} a_i \varepsilon_i \tilde{Y}_i\|_p \\
&\le \|\sum_{i=1}^{n} a_i \varepsilon_i\|_p + C_k \|\sum_{i=1}^{n} a_i \varepsilon_i Z_i\|_p \\
&\le C_k \|\sum_{i=1}^{n} a_i \varepsilon_i \mathbb{E}_Z Z_i\|_p + C_k \|\sum_{i=1}^{n} a_i \varepsilon_i Z_i\|_p \\
&\le C_k \|\sum_{i=1}^{n} a_i \varepsilon_i Z_i\|_p = C_k \|\sum_{i=1}^{n} a_i g_{i_1} \cdots g_{ik}\|_p,
\end{aligned}$$

where in the second inequality we used the contraction principle (once conditionally on $\tilde{Y}_i$'s and Z_i's) and in the third one Jensen's inequality.

References

[1] R. Adamczak. Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses. *Bull. Pol. Acad. Sci. Math.*, 53(2):221–238, 2005.

[2] R. Adamczak and R. Latała. Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails. *Ann. Inst. H. Poincar Probab. Statist.*, Volume 48, Number 4:1103–1136, 2012.

[3] S. Aida and D. Stroock. Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. *Math. Res. Lett.*, 1(1):75–86, 1994.

[4] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[5] M. A. Arcones and E. Giné. On decoupling, series expansions, and tail behavior of chaos processes. *J. Theoret. Probab.*, 6(1):101–122, 1993.

[6] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[7] D. Bakry and M. Émery. Inégalités de Sobolev pour un semi-groupe symétrique. *C. R. Acad. Sci. Paris Sér. I Math.*, 301(8):411–413, 1985.

[8] F. Barthe and E. Milman. Transference Principles for Log-Sobolev and Spectral-Gap with Applications to Conservative Spin Systems. *Available at http://arxiv.org/abs/1202.5318*, Feb. 2012.

[9] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3):383–400, 1997.

[10] S. G. Bobkov. Remarks on the growth of L^p-norms of polynomials. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 27–35. Springer, Berlin, 2000.

[11] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[12] S. G. Bobkov, F. Götze, and A. N. Tikhomirov. On concentration of empirical measures and convergence to the semi-circle law. *J. Theoret. Probab.*, 23(3):792–823, 2010.

[13] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970.

[14] C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216, 1975.

[15] C. Borell. On the Taylor series of a Wiener polynomial. *Seminar Notes on multiple stochastic integration, polynomial chaos and their integration. Case Western Reserve Univ., Cleveland*, 1984.

[16] A. N. Borodin and I. A. Ibragimov. Limit theorems for functionals of random walks. *Trudy Mat. Inst. Steklov.*, 195(Predel. Teoremy dlya Funktsional. ot Sluchain. Bluzh.):286, 1994.

[17] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.

[18] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.

[19] J. Bourgain. On the distribution of polynomials on high-dimensional convex sets. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 127–137. Springer, Berlin, 1991.

[20] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000.

[21] A. Carbery and J. Wright. Distributional and L^q norm inequalities for polynomials over convex bodies in $\mathbb{R}^n$. *Math. Res. Lett.*, 8(3):233–248, 2001.

[22] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40(4):437–451, 2012.

[23] V. H. de la Peña and E. Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York, 1999. From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond.

[24] V. H. de la Peña and S. J. Montgomery-Smith. Decoupling inequalities for the tail probabilities of multivariate U-statistics. *Ann. Probab.*, 23(2):806–816, 1995.

[25] B. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Structures Algorithms*, 41(4):469–487, 2012.

[26] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Structures Algorithms*, 40(4):452–459, 2012.

[27] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic Sobolev inequalities and transportation inequalities. *Probab. Theory Related Fields*, 133(3):409–436, 2005.

[28] E. D. Gluskin and S. Kwapień. Tail and moment estimates for sums of independent random variables with logarithmically concave tails. *Studia Math.*, 114(3):303–309, 1995.

[29] A. Guillin and A. Joulin. Measure concentration through non-Lipschitz observables and functional inequalities. *Available at http://arxiv.org/abs/1202.2341*, Feb. 2012.

[30] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000.

[31] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, 42:1079–1083, 1971.

[32] P. Hitczenko, S. J. Montgomery-Smith, and K. Oleszkiewicz. Moment inequalities for sums of certain independent symmetric random variables. *Studia Math.*, 123(1):15–42, 1997.

[33] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[34] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[35] S. Janson and A. Ruciński. The infamous upper tail. *Random Structures Algorithms*, 20(3):317–342, 2002. Probabilistic methods in combinatorial optimization.

[36] J. H. Kim and V. H. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

[37] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004.

[38] B. Klartag. Concentration of measures supported on the cube. *Available at http://arxiv.org/abs/1208.1125*, Aug. 2012.

[39] S. Kwapień. Decoupling inequalities for polynomial chaos. *Ann. Probab.*, 15(3):1062–1071, 1987.

[40] S. Kwapień and J. Szulga. Hypercontraction methods in moment inequalities for series of independent random variables in normed spaces. *Ann. Probab.*, 19(1):369–379, 1991.

[41] S. Kwapień and W. A. Woyczyński. *Random series and stochastic integrals: single and multiple.* Probability and its Applications. Birkhäuser Boston Inc., Boston, MA, 1992.

[42] R. Latała. Tail and moment estimates for some types of chaos. *Studia Math.*, 135(1):39–53, 1999.

[43] R. Latała. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.*, 34(6):2315–2331, 2006.

[44] R. Latała and R. Łochowski. Moment and tail estimates for multidimensional chaos generated by positive random variables with logarithmically concave tails. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 77–92. Birkhäuser, Basel, 2003.

[45] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[46] M. Ledoux and K. Oleszkiewicz. On measure concentration of vector-valued maps. *Bull. Pol. Acad. Sci. Math.*, 55(3):261–278, 2007.

[47] R. Łochowski. Moment and tail estimates for multidimensional chaoses generated by symmetric random variables with logarithmically concave tails. In *Approximation and probability*, volume 72 of *Banach Center Publ.*, pages 161–176. Polish Acad. Sci., Warsaw, 2006.

[48] A. Lytova and L. Pastur. On asymptotic behavior of multilinear eigenvalue statistics of random matrices. *J. Stat. Phys.*, 133(5):871–882, 2008.

[49] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[50] E. Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.*, 177(1):1–43, 2009.

[51] E. Milman. Properties of isoperimetric, functional and transport-entropy inequalities via concentration. *Probab. Theory Related Fields*, 152(3-4):475–507, 2012.

[52] F. Nazarov, M. Sodin, and A. Vol′berg. The geometric Kannan-Lovász-Simonovits lemma, dimension-free estimates for the distribution of the values of polynomials, and the distribution of the zeros of random analytic functions. *Algebra i Analiz*, 14(2):214–234, 2002.

[53] E. Nelson. The free Markoff field. *J. Funct. Anal.*, 12:211–227, 1973.

[54] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

[55] G. Pisier. Probabilistic methods in the geometry of Banach spaces. In *Probability and analysis (Varenna, 1985)*, volume 1206 of *Lecture Notes in Math.*, pages 167–241. Springer, Berlin, 1986.

[56] G. Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989.

[57] W. Schudy and M. Sviridenko. Bernstein-like Concentration and Moment Inequalities for Polynomials of Independent Random Variables: Multilinear Case. *Available at http://arxiv.org/abs/1109.5193*, Sept. 2011.

[58] W. Schudy and M. Sviridenko. Concentration and Moment Inequalities for Polynomials of Independent Random Variables. *Available at http://arxiv.org/abs/1104.4997*, Apr. 2011.

[59] A. V. Skorohod and N. P. Slobodenjuk. *Predelnye teoremy dlya sluchainykh bluzhdanii*. Izdat. "Naukova Dumka", Kiev, 1970.

[60] V. N. Sudakov and B. S. Cirel′son. Extremal properties of half-spaces for spherically invariant measures. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 41:14–24, 165, 1974. Problems in the theory of probability distributions, II.

[61] M. Talagrand. An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. *Proc. Amer. Math. Soc.*, 104(3):905–909, 1988.

[62] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[63] V. H. Vu. Concentration of non-Lipschitz functions and applications. *Random Structures Algorithms*, 20(3):262–316, 2002. Probabilistic methods in combinatorial optimization.

E-mails: `R.Adamczak@mimuw.edu.pl`, `P.Wolff@mimuw.edu.pl`